Document made available under the Patent Cooperation Treaty (PCT)

International application number: PCT/US05/010913

International filing date: 31 March 2005 (31.03.2005)

Document type: Certified copy of priority document

Document details: Country/Office: US
Number: 60/559,225
Filing date: 01 April 2004 (01.04.2004)

Date of receipt at the International Bureau: 12 August 2005 (12.08.2005)

Remark: Priority document submitted or transmitted to the International Bureau in compliance with Rule 17.1(a) or (b)

1352386

THE UNITED STATES OF AMERICA

TO ALL TO WHOM THESE PRESENTS SHALL COME:

UNITED STATES DEPARTMENT OF COMMERCE

United States Patent and Trademark Office

August 02, 2005

THIS IS TO CERTIFY THAT ANNEXED HERETO IS A TRUE COPY FROM THE RECORDS OF THE UNITED STATES PATENT AND TRADEMARK OFFICE OF THOSE PAPERS OF THE BELOW IDENTIFIED PATENT APPLICATION THAT MET THE REQUIREMENTS TO BE GRANTED A FILING DATE.

APPLICATION NUMBER: ***60/559,225***
FILING DATE: ***April 01, 2004***
RELATED PCT APPLICATION NUMBER: ***PCT/US05/10913***

UNITED STATES PATENT AND TRADEMARK OFFICE · DEPARTMENT OF COMMERCE

Certified by

Jon W. Dudas

Under Secretary of Commerce
for Intellectual Property
and Director of the United States
Patent and Trademark Office

17236 U.S.PTO
040104

22154 U.S.PTO
60/559225
040104

PTO/SB/16 (08-03)
Approved for use through 7/31/2006. OMB 0651-0032
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it displays a valid OMB control number.

PROVISIONAL APPLICATION FOR PATENT COVER SHEET

This is a request for filing a PROVISIONAL APPLICATION FOR PATENT under 37 CFR 1.53(c).

Express Mail Label No. EL 961008047 US

INVENTOR(S)

Given Name (first and middle [if any])	Family Name or Surname	Residence (City and either State or Foreign Country)
Steven	MAH	San Diego, California
Andreas	BRAUN	San Diego, California
Stefan M.	KAMMERER	San Diego, California

Additional inventors are being named on the 1 *separately numbered sheets attached hereto*

TITLE OF THE INVENTION (500 characters max)

METHODS FOR IDENTIFYING RISK OF OSTEOARTHRITIS AND TREATMENTS THEREOF

Direct all correspondence to: **CORRESPONDENCE ADDRESS**

[x] Customer Number: 25225

OR

[ ] Firm or Individual Name					
Address					
City		State		Zip	
Country		Telephone		Fax	

ENCLOSED APPLICATION PARTS (*check all that apply*)

[x] Specification *Number of Pages* 138 — [ ] CD(s), Number

[x] Drawing(s) *Number of Sheets* 1 — [x] Other (specify): Return Receipt Postcard

[x] Application Data Sheet. See 37 CFR 1.76 (4 pages)

METHOD OF PAYMENT OF FILING FEES FOR THIS PROVISIONAL APPLICATION FOR PATENT

[x] Applicant claims small entity status. See 37 CFR 1.27.

[ ] A check or money order is enclosed to cover the filing fees.

[x] The Director is hereby authorized to charge filing fees or credit any overpayment to Deposit Account Number: 03-1952

[ ] Payment by credit card. Form PTO-2038 is attached.

FILING FEE AMOUNT ($): 80.00

The invention was made by an agency of the United States Government or under a contract with an agency of the United States Government.

[x] No [ ] Yes, the name of the U.S. Government agency and the Government contract number are:

[Page 1 of 2]

Respectfully submitted,

Date April 1, 2004

SIGNATURE

TYPED OR PRINTED NAME Bruce D. Grant

REGISTRATION NO. (*if appropriate*) 47,608

TELEPHONE (858) 720-7962

Docket Number: 524593009200

USE ONLY FOR FILING A PROVISIONAL APPLICATION FOR PATENT

I hereby certify that this correspondence is being deposited with the U.S. Postal Service as Express Mail, Airbill No. EL 961008047 US, in an envelope addressed to: Mail Stop Provisional Patent Application, Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450, on the date shown below.

Dated: 4/1/04 Signature: (Deborah Wykes)

SD-189477

PROVISIONAL APPLICATION COVER SHEET
Additional Page

PTO/SB/16 (08-03)
Approved for use through 07/31/06. OMB 0651-0032
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE

Docket Number	524593009200

INVENTOR(S)/APPLICANT(S)

Given Name (first and middle [if any])	Family or Surname	Residence (City and either State or Foreign Country)
Matthew Roberts	NELSON	San Marcos, California
Rikard Henry	RENELAND	San Diego, California
Maria L.	LANGDOWN	San Diego, California

METHODS FOR IDENTIFYING RISK OF OSTEOARTHRITIS AND TREATMENTS THEREOF

Field of the Invention

[0001] The invention relates to genetic methods for identifying risk of osteoarthritis and treatments that specifically target such diseases.

Background

[0002] Osteoarthritis (OA) is a chronic disease usually affecting weight-bearing synovial joints. There are approximately 20 million Americans affected by OA and it is the leading cause of disability in the United States. In addition to extensive human suffering, OA also accounts for nearly all knee replacements and more than half of all hip replacements in the United States. Despite its prevalence, OA is poorly understood and there are few treatments available besides anti-inflammatory drugs and joint replacement.

[0003] Most commonly affecting middle-aged and older people, OA can range from very mild to very severe. It affects hands and weight-bearing joints such as knees, hips, feet and the back. Knee OA can be as disabling as any cardiovascular disease except stroke.

[0004] OA is characterized by the breakdown of cartilage in joints. Cartilage in joints cushions the ends of bones, and cartilage breakdown causes bones to rub against each other, causing pain and loss of movement. Type II collagen is the main component of cartilage, comprising 15-25% of the wet weight, approximately half the dry weight, and representing 90-95% of the total collagen content in the tissue. It forms fibrils that endow cartilage with tensile strength (Mayne, R. Arthritis Rhuem. 32:241-246 (1989)).

Summary

[0005] It has been discovered that certain polymorphic variations in human genomic DNA are associated with osteoarthritis. In particular, polymorphic variants in a locus containing a *APOL3* region in human genomic DNA have been associated with risk of osteoarthritis.

[0006] Thus, featured herein are methods for identifying a subject at risk of osteoarthritis and/or a risk of osteoarthritis in a subject, which comprise detecting the presence or absence of one or more polymorphic variations associated with osteoarthritis in or around the loci described herein in a human nucleic acid sample. In an embodiment, two or more polymorphic variations are detected in two or more regions of which one is the *APOL3* region. In certain embodiments, 3 or more, or 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19 or 20 or more polymorphic variants are detected.

[0007] Also featured are nucleic acids that include one or more polymorphic variations associated with occurrence of osteoarthritis, as well as polypeptides encoded by these nucleic acids. In addition, provided are methods for identifying candidate therapeutic molecules for treating osteoarthritis, as well as methods for treating osteoarthritis in a subject by identifying a subject at risk of osteoarthritis and treating the subject with a suitable prophylactic, treatment or therapeutic molecule.

[0008] Also provided are compositions comprising a cell from a subject having osteoarthritis or at risk of osteoarthritis and/or a *APOL3* nucleic acid, with a RNAi, siRNA, antisense DNA or RNA, or ribozyme nucleic acid designed from a *APOL3* nucleotide sequence. In an embodiment, the RNAi, siRNA, antisense DNA or RNA, or ribozyme nucleic acid is designed from a *APOL3* nucleotide sequence that includes one or more polymorphic variations associated with osteoarthritis, and in some instances, specifically interacts with such a nucleotide sequence. Further, provided are arrays of nucleic acids bound to a solid surface, in which one or more nucleic acid molecules of the array have a *APOL3* nucleotide sequence, or a fragment or substantially identical nucleic acid thereof, or a complementary nucleic acid of the foregoing. Featured also are compositions comprising a cell from a subject having osteoarthritis or at risk of osteoarthritis and/or a *APOL3* polypeptide, with an antibody that specifically binds to the polypeptide. In an embodiment, the antibody specifically binds to an epitope in the polypeptide that includes a non-synonymous amino acid modification associated with osteoarthritis (e.g., results in an amino acid substitution in the encoded polypeptide associated with osteoarthritis).

Brief Description of the Drawings

[0009] Figure 1 shows proximal SNPs in a *APOL3* region in genomic DNA. The position of each SNP in the chromosome is shown on the x-axis and the y-axis provides the negative logarithm of the p-value comparing the estimated allele to that of the control group. Also shown in the figure are exons and introns of the gene region in the approximate chromosomal positions.

Detailed Description

[0010] It has been discovered that a polymorphic variant in a locus containing a *APOL3* region is associated with occurrence of osteoarthritis in subjects. Thus, detecting genetic determinants associated with an increased risk of osteoarthritis occurrence can lead to early identification of a predisposition to osteoarthritis and early prescription of preventative measures. Also, associating a *APOL3* polymorphic variant with osteoarthritis has provided new targets for screening molecules useful in treatments of osteoarthritis.

Osteoarthritis and Sample Selection

[0011] Osteoarthritis (OA), or degenerative joint disease, is one of the oldest and most common types of arthritis. It is characterized by the breakdown of the joint's cartilage. Cartilage is the part of the joint that cushions the ends of bones, and its breakdown causes bones to rub against each other, causing pain and loss of movement. Type II collagen is the main component of cartilage, comprising 15-25% of the wet weight, approximately half the dry weight, and representing 90-95% of the total collagen content in the tissue. It forms fibrils that endow cartilage with tensile strength (Mayne, R. Arthritis Rhuem. 32:241-246 (1989)).

[0012] Most commonly affecting middle-aged and older people, OA can range from very mild to very severe. It affects hands and weight-bearing joints such as knees, hips, feet and the back. Knee OA can be as disabling as any cardiovascular disease except stroke.

[0013] Osteoarthritis affects an estimated 20.7 million Americans, mostly after age 45, with women more commonly affected than men. Physicians make a diagnosis of OA based on a physical exam and history of symptoms. X-rays are used to confirm diagnosis. Most people over 60 reflect the disease on X-ray, and about one-third have actual symptoms.

[0014] There are many factors that can cause OA. Obesity may lead to osteoarthritis of the knees. In addition, people with joint injuries due to sports, work-related activity or accidents may be at increased risk of developing OA.

[0015] Genetics has a role in the development of OA. Some people may be born with defective cartilage or with slight defects in the way that joints fit together. As a person ages, these defects may cause early cartilage breakdown in the joint or the inability to repair damaged or deteriorated cartilage in the joint.

[0016] Inclusion or exclusion of samples for an osteoarthritis pool may be based upon the following criteria: ethnicity (e.g., samples derived from an individual characterized as Caucasian); parental ethnicity (e.g., samples derived from an individual of British paternal and maternal descent); relevant phenotype information for the individual (e.g., case samples derived from individuals diagnosed with specific knee osteoarthritis (OA) and were recruited from an OA knee replacement clinic). Control samples may be selected based on relevant phenotype information for the individual (e.g., derived from individuals free of OA at several sites (knee, hand, hip etc)); and no family history of OA and/or rheumatoid arthritis. Additional phenotype information collected for both cases and controls may include age of the individual, gender, family history of OA, diagnosis with osteoarthritis (joint location of OA, date of primary diagnosis, age of individual as of primary diagnosis), knee history (current symptoms, any major knee injury, menisectomy, knee replacement surgery, age of surgery), HRT history, osteoporosis diagnosis.

[0017] Based in part upon selection criteria set forth above, individuals having osteoarthritis can be selected for genetic studies. Also, individuals having no history of osteoarthritis often are selected for genetic studies, as described hereafter.

Polymorphic Variants Associated with Osteoarthritis

[0018] A genetic analysis provided herein linked osteoarthritis with polymorphic variant nucleic acid sequences in the human genome. As used herein, the term "polymorphic site" refers to a region in a nucleic acid at which two or more alternative nucleotide sequences are observed in a significant number of nucleic acid samples from a population of individuals. A polymorphic site may be a nucleotide sequence of two or more nucleotides, an inserted nucleotide or nucleotide sequence, a deleted nucleotide or nucleotide sequence, or a microsatellite, for example. A polymorphic site that is two or more nucleotides in length may be 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15 or more, 20 or more, 30 or more, 50 or more, 75 or more, 100 or more, 500 or more, or about 1000 nucleotides in length, where all or some of the nucleotide sequences differ within the region. A polymorphic site is often one nucleotide in length, which is referred to herein as a "single nucleotide polymorphism" or a "SNP."

[0019] Where there are two, three, or four alternative nucleotide sequences at a polymorphic site, each nucleotide sequence is referred to as a "polymorphic variant" or "nucleic acid variant." Where two polymorphic variants exist, for example, the polymorphic variant represented in a minority of samples from a population is sometimes referred to as a "minor allele" and the polymorphic variant that is more prevalently represented is sometimes referred to as a "major allele." Many organisms possess a copy of each chromosome (*e.g.*, humans), and those individuals who possess two major alleles or two minor alleles are often referred to as being "homozygous" with respect to the polymorphism, and those individuals who possess one major allele and one minor allele are normally referred to as being "heterozygous" with respect to the polymorphism. Individuals who are homozygous with respect to one allele are sometimes predisposed to a different phenotype as compared to individuals who are heterozygous or homozygous with respect to another allele.

[0020] In genetic analysis that associate polymorphic variants with osteoarthritis, samples from individuals having osteoarthritis and individuals not having osteoarthritis often are allelotyped and/or genotyped. The term "allelotype" as used herein refers to a process for determining the allele frequency for a polymorphic variant in pooled DNA samples from cases and controls. By pooling DNA from each group, an allele frequency for each SNP in each group is calculated. These allele frequencies are then compared to one another. The term "genotyped" as used herein refers to a process for determining a genotype of one or more individuals, where a "genotype" is a representation of one or more polymorphic variants in a population.

[0021] A genotype or polymorphic variant may be expressed in terms of a "haplotype," which as used herein refers to two or more polymorphic variants occurring within genomic DNA in a group of individuals within a population. For example, two SNPs may exist within a gene where each SNP position includes a cytosine variation and an adenine variation. Certain individuals in a population may carry one allele (heterozygous) or two alleles (homozygous) having the gene with a cytosine at each SNP position. As the two cytosines corresponding to each SNP in the gene travel together on one or both alleles in these individuals, the individuals can be characterized as having a cytosine/cytosine haplotype with respect to the two SNPs in the gene.

[0022] As used herein, the term "phenotype" refers to a trait which can be compared between individuals, such as presence or absence of a condition, a visually observable difference in appearance between individuals, metabolic variations, physiological variations, variations in the function of biological molecules, and the like. An example of a phenotype is occurrence of osteoarthritis.

[0023] Researchers sometimes report a polymorphic variant in a database without determining whether the variant is represented in a significant fraction of a population. Because a subset of these reported polymorphic variants are not represented in a statistically significant portion of the population, some of them are sequencing errors and/or not biologically relevant. Thus, it is often not known whether a reported polymorphic variant is statistically significant or biologically relevant until the presence of the variant is detected in a population of individuals and the frequency of the variant is determined. Methods for detecting a polymorphic variant in a population are described herein, specifically in Example 2. A polymorphic variant is statistically significant and often biologically relevant if it is represented in 5% or more of a population, sometimes 10% or more, 15% or more, or 20% or more of a population, and often 25% or more, 30% or more, 35% or more, 40% or more, 45% or more, or 50% or more of a population.

[0024] A polymorphic variant may be detected on either or both strands of a double-stranded nucleic acid. Also, a polymorphic variant may be located within an intron or exon of a gene or within a portion of a regulatory region such as a promoter, a 5′ untranslated region (UTR), a 3′ UTR, and in DNA (*e.g.*, genomic DNA (gDNA) and complementary DNA (cDNA)), RNA (*e.g.*, mRNA, tRNA, and rRNA), or a polypeptide. Polymorphic variations may or may not result in detectable differences in gene expression, polypeptide structure, or polypeptide function.

[0025] It was determined that polymorphic variations associated with an increased risk of osteoarthritis existed in a *APOL3* region in SEQ ID NO: 1. In certain embodiments, a polymorphic variant at position rs132659 in the human genome was associated with an increased risk of osteoarthritis, and in a specific embodiment, a cytosine at position rs132659 was associated with an increased risk of osteoarthritis.

[0026] Polymorphic variants in and around the *APOL3* locus were tested for association with osteoarthritis. These include polymorphic variants at positions in SEQ ID NO: 1 selected from the group

consisting of 201, 425, 1095, 2201, 7879, 8395, 8461, 9503, 10304, 10695, 16300, 16444, 17591, 17988, 19116, 19358, 20300, 20669, 20891, 21451, 21978, 22785, 24248, 24770, 24844, 25066, 25096, 25309, 25344, 25529, 25537, 25554, 27963, 28134, 28356, 29648, 29986, 30217, 30267, 30315, 30585, 30724, 30897, 30931, 31080, 31246, 31373, 31463, 31467, 32188, 32288, 32520, 32594, 32657, 32677, 32764, 32784, 32830, 32872, 33121, 33348, 33952, 34184, 34361, 35026, 35192, 35600, 36033, 36289, 38869, 39629, 40530, 41621, 42379, 42802, 42865, 43644, 45051, 45828, 45829, 46257, 47286, 47427, 47963, 48013, 48229, 48282, 48376, 48404, 49900, 52699, 52897, 53414, 53487, 54112, 55492, 59766, 60307, 60701, 60952, 61401, 62379, 62870, 62879, 63499, 64284, 64408, 64760, 65230, 66127, , 6634, 66686, 66694, 67113, 67257, 67403, 67609, 68418, 68610, 69629, 70024, 70848, 71428, 71553, 71633, 71768 ,71769, 73039, 73325, 73412, 73547, 73769, 73806, 74467, 74472, 74473, 74482, 74494, 74592, 74670, 74672, 74714, 74723, 74749, 74861, 74892, 74893, 75176, 75705, 75989, 76027, 77949, 77974, 78167, 78310, 78415, 78575, 78590, 78709, 78875, 79864, 81316, 81320, 81409, 81737, 81843, 82102, 82833, 83461, 83624, 83660, 83701, 83708, 83782, 85707, 85717, 86486, 86833, 87115, 87234, 87479, 87561, 87604, 87674, 87958, 87992, 88019, 88074, 88079, 88115, 88118, 88120, 88135, 88142 ,88143, 88149, 88340, 88344, 88512, 88521, 88650, 88827, 89230, 89236, 90754, 90984, 91110, 92026, 92954, 93375, 93794, 94937, 95068, 96188, 97092 and 98812. Polymorphic variants at the following positions in SEQ ID NO: 1 in particular were associated with an increased risk of osteoarthritis: 20300, 46257, 87958, 89236, 30267, 32657, 36289, 38869, 45051, 54112, 60307, 63499, 20891, 52699, 71768, with specific embodiments directed to position 46257. In particular, the following polymorphic variants in SEQ ID NO: 1 were associated with risk of osteoarthritis: an adenine at position 20300, a thymine at position 46257, an adenine at position 89236, a guanine at position 30267, an adenine at position 32657, a cytosine at position 36289, a guanine at position 38869, a thymine at position 45051, a guanine at position 54112, an adenine at position 60307, a thymine at position 63499, a guanine at position 20891, a guanine at position 52699, and a cytosine at position 71768.

[0027] Based in part upon analyses summarized in Figure 1, a region with significant association has been identified in a locus associated with osteoarthritis. Any polymorphic variants associated with osteoarthritis in a region of significant association can be utilized for embodiments described herein. For example, polymorphic variants in a region approximately 5000 nucleotides in length and spanning chromosome positions 34828750 and 34833750 in a APOL3 locus have significant association (chromosome positions are within NCBI's Genome build 34)

Additional Polymorphic Variants Associated with Osteoarthritis

[0028] Also provided is a method for identifying polymorphic variants proximal to an incident, founder polymorphic variant associated with osteoarthritis. Thus, featured herein are methods for identifying a polymorphic variation associated with osteoarthritis that is proximal to an incident

polymorphic variation associated with osteoarthritis, which comprises identifying a polymorphic variant proximal to the incident polymorphic variant associated with osteoarthritis, where the incident polymorphic variant is in a *APOL3* nucleotide sequence. The nucleotide sequence often comprises a polynucleotide sequence selected from the group consisting of (a) a polynucleotide sequence of SEQ ID NO: 1-7; (b) a polynucleotide sequence that encodes a polypeptide having an amino acid sequence encoded by a polynucleotide sequence of SEQ ID NO: 1-7; and (c) a polynucleotide sequence that encodes a polypeptide having an amino acid sequence that is 90% or more identical to an amino acid sequence encoded by a nucleotide sequence of SEQ ID NO: 1-7 or a polynucleotide sequence 90% or more identical to the polynucleotide sequence of SEQ ID NO: 1-7. The presence or absence of an association of the proximal polymorphic variant with osteoarthritis then is determined using a known association method, such as a method described in the Examples hereafter. In an embodiment, the incident polymorphic variant is a polymorphic variant associated with osteoarthritis described herein. In another embodiment, the proximal polymorphic variant identified sometimes is a publicly disclosed polymorphic variant, which for example, sometimes is published in a publicly available database. In other embodiments, the polymorphic variant identified is not publicly disclosed and is discovered using a known method, including, but not limited to, sequencing a region surrounding the incident polymorphic variant in a group of nucleic samples. Thus, multiple polymorphic variants proximal to an incident polymorphic variant are associated with osteoarthritis using this method.

[0029] The proximal polymorphic variant often is identified in a region surrounding the incident polymorphic variant. In certain embodiments, this surrounding region is about 50 kb flanking the first polymorphic variant (*e.g.* about 50 kb 5' of the first polymorphic variant and about 50 kb 3' of the first polymorphic variant), and the region sometimes is composed of shorter flanking sequences, such as flanking sequences of about 40 kb, about 30 kb, about 25 kb, about 20 kb, about 15 kb, about 10 kb, about 7 kb, about 5 kb, or about 2 kb 5' and 3' of the incident polymorphic variant. In other embodiments, the region is composed of longer flanking sequences, such as flanking sequences of about 55 kb, about 60 kb, about 65 kb, about 70 kb, about 75 kb, about 80 kb, about 85 kb, about 90 kb, about 95 kb, or about 100 kb 5' and 3' of the incident polymorphic variant.

[0030] In certain embodiments, polymorphic variants associated with osteoarthritis are identified iteratively. For example, a first proximal polymorphic variant is associated with osteoarthritis using the methods described above and then another polymorphic variant proximal to the first proximal polymorphic variant is identified (*e.g.*, publicly disclosed or discovered) and the presence or absence of an association of one or more other polymorphic variants proximal to the first proximal polymorphic variant with osteoarthritis is determined.

[0031] The methods described herein are useful for identifying or discovering additional polymorphic variants that may be used to further characterize a gene, region or loci associated with a

condition, a disease (*e.g.*, osteoarthritis), or a disorder. For example, allelotyping or genotyping data from the additional polymorphic variants may be used to identify a functional mutation or a region of linkage disequilibrium. In certain embodiments, polymorphic variants identified or discovered within a region comprising the first polymorphic variant associated with osteoarthritis are genotyped using the genetic methods and sample selection techniques described herein, and it can be determined whether those polymorphic variants are in linkage disequilibrium with the first polymorphic variant. The size of the region in linkage disequilibrium with the first polymorphic variant also can be assessed using these genotyping methods. Thus, provided herein are methods for determining whether a polymorphic variant is in linkage disequilibrium with a first polymorphic variant associated with osteoarthritis, and such information can be used in prognosis/diagnosis methods described herein.

Isolated Nucleic Acids

[0032] Featured herein are isolated *APOL3* nucleic acid variants depicted in SEQ ID NO: 1-7, and substantially identical nucleic acids thereof. A nucleic acid variant may be represented on one or both strands in a double-stranded nucleic acid or on one chromosomal complement (heterozygous) or both chromosomal complements (homozygous).

[0033] As used herein, the term "nucleic acid" includes DNA molecules (*e.g.*, a complementary DNA (cDNA) and genomic DNA (gDNA)) and RNA molecules (*e.g.*, mRNA, rRNA, siRNA and tRNA) and analogs of DNA or RNA, for example, by use of nucleotide analogs. The nucleic acid molecule can be single-stranded and it is often double-stranded. The term "isolated or purified nucleic acid" refers to nucleic acids that are separated from other nucleic acids present in the natural source of the nucleic acid. For example, with regard to genomic DNA, the term "isolated" includes nucleic acids which are separated from the chromosome with which the genomic DNA is naturally associated. An "isolated" nucleic acid is often free of sequences which naturally flank the nucleic acid (*i.e.*, sequences located at the 5' and/or 3' ends of the nucleic acid) in the genomic DNA of the organism from which the nucleic acid is derived. For example, in various embodiments, the isolated nucleic acid molecule can contain less than about 5 kb, 4 kb, 3 kb, 2 kb, 1 kb, 0.5 kb or 0.1 kb of 5' and/or 3' nucleotide sequences which flank the nucleic acid molecule in genomic DNA of the cell from which the nucleic acid is derived. Moreover, an "isolated" nucleic acid molecule, such as a cDNA molecule, can be substantially free of other cellular material, or culture medium when produced by recombinant techniques, or substantially free of chemical precursors or other chemicals when chemically synthesized. As used herein, the term "gene" refers to a nucleotide sequence that encodes a polypeptide.

[0034] Also included herein are nucleic acid fragments. These fragments often have a nucleotide sequence identical to a nucleotide sequence of SEQ ID NO: 1-7, a nucleotide sequence substantially identical to a nucleotide sequence of SEQ ID NO: 1-7, or a nucleotide sequence that is complementary to

the foregoing. The nucleic acid fragment may be identical, substantially identical or homologous to a nucleotide sequence in an exon or an intron in a nucleotide sequence of SEQ ID NO: 1-7, and may encode a domain or part of a domain of a polypeptide. Sometimes, the fragment will comprises one or more of the polymorphic variations described herein as being associated with osteoarthritis. The nucleic acid fragment is often 50, 100, or 200 or fewer base pairs in length, and is sometimes about 300, 400, 500, 600, 700, 800, 900, 1000, 1100, 1200, 1300, 1400, 1500, 2000, 3000, 4000, 5000, 10000, 15000, or 20000 base pairs in length. A nucleic acid fragment that is complementary to a nucleotide sequence identical or substantially identical to a nucleotide sequence in SEQ ID NO: 1-7 and hybridizes to such a nucleotide sequence under stringent conditions is often referred to as a "probe." Nucleic acid fragments often include one or more polymorphic sites, or sometimes have an end that is adjacent to a polymorphic site as described hereafter.

[0035] An example of a nucleic acid fragment is an oligonucleotide. As used herein, the term "oligonucleotide" refers to a nucleic acid comprising about 8 to about 50 covalently linked nucleotides, often comprising from about 8 to about 35 nucleotides, and more often from about 10 to about 25 nucleotides. The backbone and nucleotides within an oligonucleotide may be the same as those of naturally occurring nucleic acids, or analogs or derivatives of naturally occurring nucleic acids, provided that oligonucleotides having such analogs or derivatives retain the ability to hybridize specifically to a nucleic acid comprising a targeted polymorphism. Oligonucleotides described herein may be used as hybridization probes or as components of prognostic or diagnostic assays, for example, as described herein.

[0036] Oligonucleotides are typically synthesized using standard methods and equipment, such as the ABI™3900 High Throughput DNA Synthesizer and the EXPEDITE™ 8909 Nucleic Acid Synthesizer, both of which are available from Applied Biosystems (Foster City, CA). Analogs and derivatives are exemplified in U.S. Pat. Nos. 4,469,863; 5,536,821; 5,541,306; 5,637,683; 5,637,684; 5,700,922; 5,717,083; 5,719,262; 5,739,308; 5,773,601; 5,886,165; 5,929,226; 5,977,296; 6,140,482; WO 00/56746; WO 01/14398, and related publications. Methods for synthesizing oligonucleotides comprising such analogs or derivatives are disclosed, for example, in the patent publications cited above and in U.S. Pat. Nos. 5,614,622; 5,739,314; 5,955,599; 5,962,674; 6,117,992; in WO 00/75372; and in related publications.

[0037] Oligonucleotides may also be linked to a second moiety. The second moiety may be an additional nucleotide sequence such as a tail sequence (*e.g.*, a polyadenosine tail), an adapter sequence (*e.g.*, phage M13 universal tail sequence), and others. Alternatively, the second moiety may be a non-nucleotide moiety such as a moiety which facilitates linkage to a solid support or a label to facilitate detection of the oligonucleotide. Such labels include, without limitation, a radioactive label, a

fluorescent label, a chemiluminescent label, a paramagnetic label, and the like. The second moiety may be attached to any position of the oligonucleotide, provided the oligonucleotide can hybridize to the nucleic acid comprising the polymorphism.

Uses for Nucleic Acid Sequence

[0038] Nucleic acid coding sequences may be used for diagnostic purposes for detection and control of polypeptide expression. Also, included herein are oligonucleotide sequences such as antisense RNA, small-interfering RNA (siRNA) and DNA molecules and ribozymes that function to inhibit translation of a polypeptide. Antisense techniques and RNA interference techniques are known in the art and are described herein.

[0039] Ribozymes are enzymatic RNA molecules capable of catalyzing the specific cleavage of RNA. The mechanism of ribozyme action involves sequence specific hybridization of the ribozyme molecule to complementary target RNA, followed by endonucleolytic cleavage. For example, hammerhead motif ribozyme molecules may be engineered that specifically and efficiently catalyze endonucleolytic cleavage of RNA sequences corresponding to or complementary to *APOL3* nucleotide sequences. Specific ribozyme cleavage sites within any potential RNA target are initially identified by scanning the target molecule for ribozyme cleavage sites which include the following sequences, GUA, GUU and GUC. Once identified, short RNA sequences of between fifteen (15) and twenty (20) ribonucleotides corresponding to the region of the target gene containing the cleavage site may be evaluated for predicted structural features such as secondary structure that may render the oligonucleotide sequence unsuitable. The suitability of candidate targets may also be evaluated by testing their accessibility to hybridization with complementary oligonucleotides, using ribonuclease protection assays.

[0040] Antisense RNA and DNA molecules, siRNA and ribozymes may be prepared by any method known in the art for the synthesis of RNA molecules. These include techniques for chemically synthesizing oligodeoxyribonucleotides well known in the art such as solid phase phosphoramidite chemical synthesis. Alternatively, RNA molecules may be generated by *in vitro* and *in vivo* transcription of DNA sequences encoding the antisense RNA molecule. Such DNA sequences may be incorporated into a wide variety of vectors which incorporate suitable RNA polymerase promoters such as the T7 or SP6 polymerase promoters. Alternatively, antisense cDNA constructs that synthesize antisense RNA constitutively or inducibly, depending on the promoter used, can be introduced stably into cell lines.

[0041] DNA encoding a polypeptide also may have a number of uses for the diagnosis of diseases, including osteoarthritis, resulting from aberrant expression of a target gene described herein. For example, the nucleic acid sequence may be used in hybridization assays of biopsies or autopsies to diagnose abnormalities of expression or function (*e.g.*, Southern or Northern blot analysis, in situ hybridization assays).

[0042] In addition, the expression of a polypeptide during embryonic development may also be determined using nucleic acid encoding the polypeptide. As addressed, *infra*, production of functionally impaired polypeptide is the cause of various disease states, such as osteoarthritis. *In situ* hybridizations using polypeptide as a probe may be employed to predict problems related to osteoarthritis. Further, as indicated, *infra*, administration of human active polypeptide, recombinantly produced as described herein, may be used to treat disease states related to functionally impaired polypeptide. Alternatively, gene therapy approaches may be employed to remedy deficiencies of functional polypeptide or to replace or compete with dysfunctional polypeptide.

Expression Vectors, Host Cells, and Genetically Engineered Cells

[0043] Provided herein are nucleic acid vectors, often expression vectors, which contain a *APOL3* nucleotide sequence, or a substantially identical sequence thereof. As used herein, the term "vector" refers to a nucleic acid molecule capable of transporting another nucleic acid to which it has been linked and can include a plasmid, cosmid, or viral vector. The vector can be capable of autonomous replication or it can integrate into a host DNA. Viral vectors may include replication defective retroviruses, adenoviruses and adeno-associated viruses for example.

[0044] A vector can include a *APOL3* nucleotide sequence in a form suitable for expression of an encoded target polypeptide or target nucleic acid in a host cell. A "target polypeptide" is a polypeptide encoded by a *APOL3* nucleotide sequence, or a substantially identical nucleotide sequence thereof. The recombinant expression vector typically includes one or more regulatory sequences operatively linked to the nucleic acid sequence to be expressed. The term "regulatory sequence" includes promoters, enhancers and other expression control elements (*e.g.*, polyadenylation signals). Regulatory sequences include those that direct constitutive expression of a nucleotide sequence, as well as tissue-specific regulatory and/or inducible sequences. The design of the expression vector can depend on such factors as the choice of the host cell to be transformed, the level of expression of polypeptide desired, and the like. Expression vectors can be introduced into host cells to produce target polypeptides, including fusion polypeptides.

[0045] Recombinant expression vectors can be designed for expression of target polypeptides in prokaryotic or eukaryotic cells. For example, target polypeptides can be expressed in *E. coli*, insect cells (*e.g.*, using baculovirus expression vectors), yeast cells, or mammalian cells. Suitable host cells are discussed further in Goeddel, *Gene Expression Technology: Methods in Enzymology 185*, Academic Press, San Diego, CA (1990). Alternatively, the recombinant expression vector can be transcribed and translated *in vitro*, for example using T7 promoter regulatory sequences and T7 polymerase.

[0046] Expression of polypeptides in prokaryotes is most often carried out in *E. coli* with vectors containing constitutive or inducible promoters directing the expression of either fusion or non-fusion

polypeptides. Fusion vectors add a number of amino acids to a polypeptide encoded therein, usually to the amino terminus of the recombinant polypeptide. Such fusion vectors typically serve three purposes: 1) to increase expression of recombinant polypeptide; 2) to increase the solubility of the recombinant polypeptide; and 3) to aid in the purification of the recombinant polypeptide by acting as a ligand in affinity purification. Often, a proteolytic cleavage site is introduced at the junction of the fusion moiety and the recombinant polypeptide to enable separation of the recombinant polypeptide from the fusion moiety subsequent to purification of the fusion polypeptide. Such enzymes, and their cognate recognition sequences, include Factor Xa, thrombin and enterokinase. Typical fusion expression vectors include pGEX (Pharmacia Biotech Inc; Smith & Johnson, *Gene 67:* 31-40 (1988)), pMAL (New England Biolabs, Beverly, MA) and pRIT5 (Pharmacia, Piscataway, NJ) which fuse glutathione S-transferase (GST), maltose E binding polypeptide, or polypeptide A, respectively, to the target recombinant polypeptide.

[0047] Purified fusion polypeptides can be used in screening assays and to generate antibodies specific for target polypeptides. In a therapeutic embodiment, fusion polypeptide expressed in a retroviral expression vector is used to infect bone marrow cells that are subsequently transplanted into irradiated recipients. The pathology of the subject recipient is then examined after sufficient time has passed (*e.g.*, six (6) weeks).

[0048] Expressing the polypeptide in host bacteria with an impaired capacity to proteolytically cleave the recombinant polypeptide is often used to maximize recombinant polypeptide expression (Gottesman, S., *Gene Expression Technology: Methods in Enzymology, Academic Press, San Diego, California 185:* 119-128 (1990)). Another strategy is to alter the nucleotide sequence of the nucleic acid to be inserted into an expression vector so that the individual codons for each amino acid are those preferentially utilized in *E. coli* (Wada *et al., Nucleic Acids Res. 20:* 2111-2118 (1992)). Such alteration of nucleotide sequences can be carried out by standard DNA synthesis techniques.

[0049] When used in mammalian cells, the expression vector's control functions are often provided by viral regulatory elements. For example, commonly used promoters are derived from polyoma, Adenovirus 2, cytomegalovirus and Simian Virus 40. Recombinant mammalian expression vectors are often capable of directing expression of the nucleic acid in a particular cell type (*e.g.*, tissue-specific regulatory elements are used to express the nucleic acid). Non-limiting examples of suitable tissue-specific promoters include an albumin promoter (liver-specific; Pinkert *et al., Genes Dev. 1:* 268-277 (1987)), lymphoid-specific promoters (Calame & Eaton, *Adv. Immunol. 43:* 235-275 (1988)), promoters of T cell receptors (Winoto & Baltimore, *EMBO J. 8:* 729-733 (1989)) promoters of immunoglobulins (Banerji *et al., Cell 33:* 729-740 (1983); Queen & Baltimore, *Cell 33:* 741-748 (1983)), neuron-specific promoters (*e.g.*, the neurofilament promoter; Byrne & Ruddle, *Proc. Natl. Acad. Sci. USA 86:* 5473-5477 (1989)), pancreas-specific promoters (Edlund *et al., Science 230:* 912-916 (1985)), and mammary

gland-specific promoters (*e.g.*, milk whey promoter; U.S. Patent No. 4,873,316 and European Application Publication No. 264,166). Developmentally-regulated promoters are sometimes utilized, for example, the murine hox promoters (Kessel & Gruss, *Science 249:* 374-379 (1990)) and the α-fetopolypeptide promoter (Campes & Tilghman, *Genes Dev. 3:* 537-546 (1989)).

[0050] A *APOL3* nucleic acid also may be cloned into an expression vector in an antisense orientation. Regulatory sequences (*e.g.*, viral promoters and/or enhancers) operatively linked to a *APOL3* nucleic acid cloned in the antisense orientation can be chosen for directing constitutive, tissue specific or cell type specific expression of antisense RNA in a variety of cell types. Antisense expression vectors can be in the form of a recombinant plasmid, phagemid or attenuated virus. For a discussion of the regulation of gene expression using antisense genes *see, e.g.,* Weintraub *et al.,* Antisense RNA as a molecular tool for genetic analysis, *Reviews - Trends in Genetics*, Vol. 1(1) (1986).

[0051] Also provided herein are host cells that include a *APOL3* nucleotide sequence within a recombinant expression vector or a fragment of such a nucleotide sequence which facilitate homologous recombination into a specific site of the host cell genome. The terms "host cell" and "recombinant host cell" are used interchangeably herein. Such terms refer not only to the particular subject cell but rather also to the progeny or potential progeny of such a cell. Because certain modifications may occur in succeeding generations due to either mutation or environmental influences, such progeny may not, in fact, be identical to the parent cell, but are still included within the scope of the term as used herein. A host cell can be any prokaryotic or eukaryotic cell. For example, a target polypeptide can be expressed in bacterial cells such as *E. coli*, insect cells, yeast or mammalian cells (such as Chinese hamster ovary cells (CHO) or COS cells). Other suitable host cells are known to those skilled in the art.

[0052] Vectors can be introduced into host cells via conventional transformation or transfection techniques. As used herein, the terms "transformation" and "transfection" are intended to refer to a variety of art-recognized techniques for introducing foreign nucleic acid (*e.g.*, DNA) into a host cell, including calcium phosphate or calcium chloride co-precipitation, transduction/infection, DEAE-dextran-mediated transfection, lipofection, or electroporation.

[0053] A host cell provided herein can be used to produce (*i.e.*, express) a target polypeptide or a substantially identical polypeptide thereof. Accordingly, further provided are methods for producing a target polypeptide using host cells described herein. In one embodiment, the method includes culturing host cells into which a recombinant expression vector encoding a target polypeptide has been introduced in a suitable medium such that a target polypeptide is produced. In another embodiment, the method further includes isolating a target polypeptide from the medium or the host cell.

[0054] Also provided are cells or purified preparations of cells which include a *APOL3* transgene, or which otherwise misexpress target polypeptide. Cell preparations can consist of human or non-human

cells, *e.g.*, rodent cells, *e.g.*, mouse or rat cells, rabbit cells, or pig cells. In preferred embodiments, the cell or cells include a *APOL3* transgene (*e.g.*, a heterologous form of a *APOL3* gene, such as a human gene expressed in non-human cells). The transgene can be misexpressed, *e.g.*, overexpressed or underexpressed. In other preferred embodiments, the cell or cells include a gene which misexpress an endogenous target polypeptide (*e.g.*, expression of a gene is disrupted, also known as a knockout). Such cells can serve as a model for studying disorders which are related to mutated or mis-expressed alleles or for use in drug screening. Also provided are human cells (*e.g.*, a hematopoietic stem cells) transfected with a *APOL3* nucleic acid.

[0055] Also provided are cells or a purified preparation thereof (*e.g.*, human cells) in which an endogenous *APOL3* nucleic acid is under the control of a regulatory sequence that does not normally control the expression of the endogenous gene. The expression characteristics of an endogenous gene within a cell (*e.g.*, a cell line or microorganism) can be modified by inserting a heterologous DNA regulatory element into the genome of the cell such that the inserted regulatory element is operably linked to the corresponding endogenous gene. For example, an endogenous corresponding gene (*e.g.*, a gene which is "transcriptionally silent," not normally expressed, or expressed only at very low levels) may be activated by inserting a regulatory element which is capable of promoting the expression of a normally expressed gene product in that cell. Techniques such as targeted homologous recombinations, can be used to insert the heterologous DNA as described in, *e.g.*, Chappel, US 5,272,071; WO 91/06667, published on May 16, 1991.

<u>Transgenic Animals</u>

[0056] Non-human transgenic animals that express a heterologous target polypeptide (*e.g.*, expressed from a *APOL3* nucleic acid or substantially identical sequence thereof) can be generated. Such animals are useful for studying the function and/or activity of a target polypeptide and for identifying and/or evaluating modulators of the activity of *APOL3* nucleic acids and encoded polypeptides. As used herein, a "transgenic animal" is a non-human animal such as a mammal (*e.g.*, a non-human primate such as chimpanzee, baboon, or macaque; an ungulate such as an equine, bovine, or caprine; or a rodent such as a rat, a mouse, or an Israeli sand rat), a bird (*e.g.*, a chicken or a turkey), an amphibian (*e.g.*, a frog, salamander, or newt), or an insect (*e.g.*, *Drosophila melanogaster*), in which one or more of the cells of the animal includes a transgene. A transgene is exogenous DNA or a rearrangement (*e.g.*, a deletion of endogenous chromosomal DNA) that is often integrated into or occurs in the genome of cells in a transgenic animal. A transgene can direct expression of an encoded gene product in one or more cell types or tissues of the transgenic animal, and other transgenes can reduce expression (*e.g.*, a knockout). Thus, a transgenic animal can be one in which an endogenous nucleic acid homologous to a *APOL3* nucleic acid has been altered by homologous recombination between the endogenous gene and an

exogenous DNA molecule introduced into a cell of the animal (*e.g.*, an embryonic cell of the animal) prior to development of the animal.

[0057] Intronic sequences and polyadenylation signals can also be included in the transgene to increase expression efficiency of the transgene. One or more tissue-specific regulatory sequences can be operably linked to a *APOL3* nucleotide sequence to direct expression of an encoded polypeptide to particular cells. A transgenic founder animal can be identified based upon the presence of a *APOL3* nucleotide sequence in its genome and/or expression of encoded mRNA in tissues or cells of the animals. A transgenic founder animal can then be used to breed additional animals carrying the transgene. Moreover, transgenic animals carrying a *APOL3* nucleotide sequence can further be bred to other transgenic animals carrying other transgenes.

[0058] Target polypeptides can be expressed in transgenic animals or plants by introducing, for example, a *APOL3* nucleic acid into the genome of an animal that encodes the target polypeptide. In preferred embodiments the nucleic acid is placed under the control of a tissue specific promoter, *e.g.*, a milk or egg specific promoter, and recovered from the milk or eggs produced by the animal. Also included is a population of cells from a transgenic animal.

Target Polypeptides

[0059] Also featured herein are isolated target polypeptides, which are encoded by a *APOL3* nucleotide sequence (e.g., SEQ ID NO: 1-7), or a substantially identical nucleotide sequence thereof. Examples of *APOL3* polypeptides are set forth in SEQ ID NO: 8-10. The term "polypeptide" as used herein includes proteins and peptides. An "isolated" or "purified" polypeptide or protein is substantially free of cellular material or other contaminating proteins from the cell or tissue source from which the protein is derived, or substantially free from chemical precursors or other chemicals when chemically synthesized. In one embodiment, the language "substantially free" means preparation of a target polypeptide having less than about 30%, 20%, 10% and more preferably 5% (by dry weight), of non-target polypeptide (also referred to herein as a "contaminating protein"), or of chemical precursors or non-target chemicals. When the target polypeptide or a biologically active portion thereof is recombinantly produced, it is also preferably substantially free of culture medium, specifically, where culture medium represents less than about 20%, sometimes less than about 10%, and often less than about 5% of the volume of the polypeptide preparation. Isolated or purified target polypeptide preparations are sometimes 0.01 milligrams or more or 0.1 milligrams or more, and often 1.0 milligrams or more and 10 milligrams or more in dry weight. In certain embodiments, the *APOL3* polypeptide or polypeptide fragment has *APOL3* biological activity, for example, apolipoprotein activity.

[0060] Further included herein are target polypeptide fragments. The polypeptide fragment may be a domain or part of a domain of a target polypeptide. The polypeptide fragment may have increased,

decreased or unexpected biological activity. The polypeptide fragment is often 50 or fewer, 100 or fewer, or 200 or fewer amino acids in length, and is sometimes 300, 400, 500, 600, 700, or 900 or fewer amino acids in length. The polypeptide fragment sometimes is amino acids 90-396; amino acids 19-325; or amino acids 1-196.

[0061] Substantially identical target polypeptides may depart from the amino acid sequences of target polypeptides in different manners. For example, conservative amino acid modifications may be introduced at one or more positions in the amino acid sequences of target polypeptides. A "conservative amino acid substitution" is one in which the amino acid is replaced by another amino acid having a similar structure and/or chemical function. Families of amino acid residues having similar structures and functions are well known. These families include amino acids with basic side chains (*e.g.*, lysine, arginine, histidine), acidic side chains (*e.g.*, aspartic acid, glutamic acid), uncharged polar side chains (*e.g.*, glycine, asparagine, glutamine, serine, threonine, tyrosine, cysteine), nonpolar side chains (*e.g.*, alanine, valine, leucine, isoleucine, proline, phenylalanine, methionine, tryptophan), beta-branched side chains (*e.g.*, threonine, valine, isoleucine) and aromatic side chains (*e.g.*, tyrosine, phenylalanine, tryptophan, histidine). Also, essential and non-essential amino acids may be replaced. A "non-essential" amino acid is one that can be altered without abolishing or substantially altering the biological function of a target polypeptide, whereas altering an "essential" amino acid abolishes or substantially alters the biological function of a target polypeptide. Amino acids that are conserved among target polypeptides are typically essential amino acids. In certain embodiments, the polypeptide includes one or more non-synonymous polymorphic variants associated with osteoarthritis.

[0062] Also, target polypeptides may exist as chimeric or fusion polypeptides. As used herein, a target "chimeric polypeptide" or target "fusion polypeptide" includes a target polypeptide linked to a non-target polypeptide. A "non-target polypeptide" refers to a polypeptide having an amino acid sequence corresponding to a polypeptide which is not substantially identical to the target polypeptide, which includes, for example, a polypeptide that is different from the target polypeptide and derived from the same or a different organism. The target polypeptide in the fusion polypeptide can correspond to an entire or nearly entire target polypeptide or a fragment thereof. The non-target polypeptide can be fused to the N-terminus or C-terminus of the target polypeptide.

[0063] Fusion polypeptides can include a moiety having high affinity for a ligand. For example, the fusion polypeptide can be a GST-target fusion polypeptide in which the target sequences are fused to the C-terminus of the GST sequences, or a polyhistidine-target fusion polypeptide in which the target polypeptide is fused at the N- or C-terminus to a string of histidine residues. Such fusion polypeptides can facilitate purification of recombinant target polypeptide. Expression vectors are commercially available that already encode a fusion moiety (*e.g.*, a GST polypeptide), and a nucleotide sequence in SEQ ID NO: 1-7, or a substantially identical nucleotide sequence thereof, can be cloned into an

expression vector such that the fusion moiety is linked in-frame to the target polypeptide. Further, the fusion polypeptide can be a target polypeptide containing a heterologous signal sequence at its N-terminus. In certain host cells (*e.g.*, mammalian host cells), expression, secretion, cellular internalization, and cellular localization of a target polypeptide can be increased through use of a heterologous signal sequence. Fusion polypeptides can also include all or a part of a serum polypeptide (*e.g.*, an IgG constant region or human serum albumin).

[0064] Target polypeptides can be incorporated into pharmaceutical compositions and administered to a subject *in vivo*. Administration of these target polypeptides can be used to affect the bioavailability of a substrate of the target polypeptide and may effectively increase target polypeptide biological activity in a cell. Target fusion polypeptides may be useful therapeutically for the treatment of disorders caused by, for example, (i) aberrant modification or mutation of a gene encoding a target polypeptide; (ii) mis-regulation of the gene encoding the target polypeptide; and (iii) aberrant post-translational modification of a target polypeptide. Also, target polypeptides can be used as immunogens to produce anti-target antibodies in a subject, to purify target polypeptide ligands or binding partners, and in screening assays to identify molecules which inhibit or enhance the interaction of a target polypeptide with a substrate.

[0065] In addition, polypeptides can be chemically synthesized using techniques known in the art (See, *e.g.*, Creighton, 1983 Proteins. New York, N.Y.: W. H. Freeman and Company; and Hunkapiller et al., (1984) Nature July 12 -18;310(5973):105-11). For example, a relative short fragment can be synthesized by use of a peptide synthesizer. Furthermore, if desired, non-classical amino acids or chemical amino acid analogs can be introduced as a substitution or addition into the fragment sequence. Non-classical amino acids include, but are not limited to, to the D-isomers of the common amino acids, 2,4-diaminobutyric acid, a-amino isobutyric acid, 4-aminobutyric acid, Abu, 2-amino butyric acid, g-Abu, e-Ahx, 6-amino hexanoic acid, Aib, 2-amino isobutyric acid, 3-amino propionic acid, ornithine, norleucine, norvaline, hydroxyproline, sarcosine, citrulline, homocitrulline, cysteic acid, t-butylglycine, t-butylalanine, phenylglycine, cyclohexylalanine, b-alanine, fluoroamino acids, designer amino acids such as b-methyl amino acids, Ca-methyl amino acids, Na-methyl amino acids, and amino acid analogs in general. Furthermore, the amino acid can be D (dextrorotary) or L (levorotary).

[0066] Polypeptides and polypeptide fragments sometimes are differentially modified during or after translation, *e.g.*, by glycosylation, acetylation, phosphorylation, amidation, derivatization by known protecting/blocking groups, proteolytic cleavage, linkage to an antibody molecule or other cellular ligand, etc. Any of numerous chemical modifications may be carried out by known techniques, including but not limited, to specific chemical cleavage by cyanogen bromide, trypsin, chymotrypsin, papain, V8 protease, NaBH4; acetylation, formylation, oxidation, reduction; metabolic synthesis in the presence of tunicamycin; and the like. Additional post-translational modifications include, for example, N-linked or O-linked carbohydrate chains, processing of N-terminal or C-terminal ends), attachment of chemical

moieties to the amino acid backbone, chemical modifications of N-linked or O-linked carbohydrate chains, and addition or deletion of an N-terminal methionine residue as a result of prokaryotic host cell expression. The polypeptide fragments may also be modified with a detectable label, such as an enzymatic, fluorescent, isotopic or affinity label to allow for detection and isolation of the polypeptide.

[0067] Also provided are chemically modified derivatives of polypeptides that can provide additional advantages such as increased solubility, stability and circulating time of the polypeptide, or decreased immunogenicity (*see e.g.*, U.S. Pat. No: 4,179,337. The chemical moieties for derivitization may be selected from water soluble polymers such as polyethylene glycol, ethylene glycol/propylene glycol copolymers, carboxymethylcellulose, dextran, polyvinyl alcohol and the like. The polypeptides may be modified at random positions within the molecule, or at predetermined positions within the molecule and may include one, two, three or more attached chemical moieties.

[0068] The polymer may be of any molecular weight, and may be branched or unbranched. For polyethylene glycol, the preferred molecular weight is between about 1 kDa and about 100 kDa (the term "about" indicating that in preparations of polyethylene glycol, some molecules will weigh more, some less, than the stated molecular weight) for ease in handling and manufacturing. Other sizes may be used, depending on the desired therapeutic profile (e.g., the duration of sustained release desired, the effects, if any on biological activity, the ease in handling, the degree or lack of antigenicity and other known effects of the polyethylene glycol to a therapeutic protein or analog).

[0069] The polymers should be attached to the polypeptide with consideration of effects on functional or antigenic domains of the polypeptide. There are a number of attachment methods available to those skilled in the art (*e.g.*, EP 0 401 384 (coupling PEG to G-CSF) and Malik et al. (1992) Exp Hematol. September;20(8):1028-35 (pegylation of GM-CSF using tresyl chloride)). For example, polyethylene glycol may be covalently bound through amino acid residues via a reactive group, such as a free amino or carboxyl group. Reactive groups are those to which an activated polyethylene glycol molecule may be bound. The amino acid residues having a free amino group may include lysine residues and the N-terminal amino acid residues; those having a free carboxyl group may include aspartic acid residues, glutamic acid residues and the C-terminal amino acid residue. Sulfhydryl groups may also be used as a reactive group for attaching the polyethylene glycol molecules. For therapeutic purposes, the attachment sometimes is at an amino group, such as attachment at the N-terminus or lysine group.

[0070] Proteins can be chemically modified at the N-terminus. Using polyethylene glycol as an illustration of such a composition, one may select from a variety of polyethylene glycol molecules (by molecular weight, branching, and the like), the proportion of polyethylene glycol molecules to protein (polypeptide) molecules in the reaction mix, the type of pegylation reaction to be performed, and the method of obtaining the selected N-terminally pegylated protein. The method of obtaining the N-terminally pegylated preparation (i.e., separating this moiety from other monopegylated moieties if

necessary) may be by purification of the N-terminally pegylated material from a population of pegylated protein molecules. Selective proteins chemically modified at the N-terminus may be accomplished by reductive alkylation, which exploits differential reactivity of different types of primary amino groups (lysine versus the N-terminal) available for derivatization in a particular protein. Under the appropriate reaction conditions, substantially selective derivatization of the protein at the N-terminus with a carbonyl group containing polymer is achieved.

Substantially Identical Nucleic Acids and Polypeptides

[0071] Nucleotide sequences and polypeptide sequences that are substantially identical to a *APOL3* nucleotide sequence and the target polypeptide sequences encoded by those nucleotide sequences, respectively, are included herein. The term "substantially identical" as used herein refers to two or more nucleic acids or polypeptides sharing one or more identical nucleotide sequences or polypeptide sequences, respectively. Included are nucleotide sequences or polypeptide sequences that are 55% or more, 60% or more, 65% or more, 70% or more, 75% or more, 80% or more, 85% or more, 90% or more, 95% or more (each often within a 1%, 2%, 3% or 4% variability) identical to a *APOL3* nucleotide sequence or the encoded target polypeptide amino acid sequences. One test for determining whether two nucleic acids are substantially identical is to determine the percent of identical nucleotide sequences or polypeptide sequences shared between the nucleic acids or polypeptides.

[0072] Calculations of sequence identity are often performed as follows. Sequences are aligned for optimal comparison purposes (*e.g.*, gaps can be introduced in one or both of a first and a second amino acid or nucleic acid sequence for optimal alignment and non-homologous sequences can be disregarded for comparison purposes). The length of a reference sequence aligned for comparison purposes is sometimes 30% or more, 40% or more, 50% or more, often 60% or more, and more often 70% or more, 80% or more, 90% or more, or 100% of the length of the reference sequence. The nucleotides or amino acids at corresponding nucleotide or polypeptide positions, respectively, are then compared among the two sequences. When a position in the first sequence is occupied by the same nucleotide or amino acid as the corresponding position in the second sequence, the nucleotides or amino acids are deemed to be identical at that position. The percent identity between the two sequences is a function of the number of identical positions shared by the sequences, taking into account the number of gaps, and the length of each gap, introduced for optimal alignment of the two sequences.

[0073] Comparison of sequences and determination of percent identity between two sequences can be accomplished using a mathematical algorithm. Percent identity between two amino acid or nucleotide sequences can be determined using the algorithm of Meyers & Miller, *CABIOS 4:* 11-17 (1989), which has been incorporated into the ALIGN program (version 2.0), using a PAM120 weight residue table, a gap length penalty of 12 and a gap penalty of 4. Also, percent identity between two amino acid

sequences can be determined using the Needleman & Wunsch, *J. Mol. Biol. 48:* 444-453 (1970) algorithm which has been incorporated into the GAP program in the GCG software package (available at the http address www.gcg.com), using either a Blossum 62 matrix or a PAM250 matrix, and a gap weight of 16, 14, 12, 10, 8, 6, or 4 and a length weight of 1, 2, 3, 4, 5, or 6. Percent identity between two nucleotide sequences can be determined using the GAP program in the GCG software package (available at http address www.gcg.com), using a NWSgapdna.CMP matrix and a gap weight of 40, 50, 60, 70, or 80 and a length weight of 1, 2, 3, 4, 5, or 6. A set of parameters often used is a Blossum 62 scoring matrix with a gap open penalty of 12, a gap extend penalty of 4, and a frameshift gap penalty of 5.

[0074] Another manner for determining if two nucleic acids are substantially identical is to assess whether a polynucleotide homologous to one nucleic acid will hybridize to the other nucleic acid under stringent conditions. As use herein, the term "stringent conditions" refers to conditions for hybridization and washing. Stringent conditions are known to those skilled in the art and can be found in *Current Protocols in Molecular Biology*, John Wiley & Sons, N.Y. , 6.3.1-6.3.6 (1989). Aqueous and non-aqueous methods are described in that reference and either can be used. An example of stringent hybridization conditions is hybridization in 6X sodium chloride/sodium citrate (SSC) at about 45°C, followed by one or more washes in 0.2X SSC, 0.1% SDS at 50°C. Another example of stringent hybridization conditions are hybridization in 6X sodium chloride/sodium citrate (SSC) at about 45°C, followed by one or more washes in 0.2X SSC, 0.1% SDS at 55°C. A further example of stringent hybridization conditions is hybridization in 6X sodium chloride/sodium citrate (SSC) at about 45°C, followed by one or more washes in 0.2X SSC, 0.1% SDS at 60°C. Often, stringent hybridization conditions are hybridization in 6X sodium chloride/sodium citrate (SSC) at about 45°C, followed by one or more washes in 0.2X SSC, 0.1% SDS at 65°C. More often, stringency conditions are 0.5M sodium phosphate, 7% SDS at 65°C, followed by one or more washes at 0.2X SSC, 1% SDS at 65°C.

[0075] An example of a substantially identical nucleotide sequence to a nucleotide sequence in SEQ ID NO: 1-7 is one that has a different nucleotide sequence but still encodes the same polypeptide sequence encoded by the nucleotide sequence in SEQ ID NO: 1-7. Another example is a nucleotide sequence that encodes a polypeptide having a polypeptide sequence that is more than 70% or more identical to, sometimes more than 75% or more, 80% or more, or 85% or more identical to, and often more than 90% or more and 95% or more identical to a polypeptide sequence encoded by a nucleotide sequence in SEQ ID NO: 1-7.

[0076] Nucleotide sequences in SEQ ID NO: 1-7 and amino acid sequences of encoded polypeptides can be used as "query sequences" to perform a search against public databases to identify other family members or related sequences, for example. Such searches can be performed using the NBLAST and XBLAST programs (version 2.0) of Altschul *et al., J. Mol. Biol. 215:* 403-10 (1990). BLAST nucleotide

searches can be performed with the NBLAST program, score = 100, wordlength = 12 to obtain nucleotide sequences homologous to nucleotide sequences in SEQ ID NO: 1-7. BLAST polypeptide searches can be performed with the XBLAST program, score = 50, wordlength = 3 to obtain amino acid sequences homologous to polypeptides encoded by the nucleotide sequences of SEQ ID NO: 1-7. To obtain gapped alignments for comparison purposes, Gapped BLAST can be utilized as described in Altschul *et al., Nucleic Acids Res. 25(17):* 3389-3402 (1997). When utilizing BLAST and Gapped BLAST programs, default parameters of the respective programs (*e.g.*, XBLAST and NBLAST) can be used (*see* the http address www.ncbi.nlm.nih.gov).

[0077] A nucleic acid that is substantially identical to a nucleotide sequence in SEQ ID NO: 1-7 may include polymorphic sites at positions equivalent to those described herein when the sequences are aligned. For example, using the alignment procedures described herein, SNPs in a sequence substantially identical to a sequence in SEQ ID NO: 1-7 can be identified at nucleotide positions that match (*i.e.*, align) with nucleotides at SNP positions in each nucleotide sequence in SEQ ID NO: 1-7. Also, where a polymorphic variation results in an insertion or deletion, insertion or deletion of a nucleotide sequence from a reference sequence can change the relative positions of other polymorphic sites in the nucleotide sequence.

[0078] Substantially identical nucleotide and polypeptide sequences include those that are naturally occurring, such as allelic variants (same locus), splice variants, homologs (different locus), and orthologs (different organism) or can be non-naturally occurring. Non-naturally occurring variants can be generated by mutagenesis techniques, including those applied to polynucleotides, cells, or organisms. The variants can contain nucleotide substitutions, deletions, inversions and insertions. Variation can occur in either or both the coding and non-coding regions. The variations can produce both conservative and non-conservative amino acid substitutions (as compared in the encoded product). Orthologs, homologs, allelic variants, and splice variants can be identified using methods known in the art. These variants normally comprise a nucleotide sequence encoding a polypeptide that is 50% or more, about 55% or more, often about 70-75% or more or about 80-85% or more, and sometimes about 90-95% or more identical to the amino acid sequences of target polypeptides or a fragment thereof. Such nucleic acid molecules can readily be identified as being able to hybridize under stringent conditions to a nucleotide sequence in SEQ ID NO: 1 or a fragment of this sequence. Nucleic acid molecules corresponding to orthologs, homologs, and allelic variants of a nucleotide sequence in SEQ ID NO: 1-7 can further be identified by mapping the sequence to the same chromosome or locus as the nucleotide sequence in SEQ ID NO: 1-7.

[0079] Also, substantially identical nucleotide sequences may include codons that are altered with respect to the naturally occurring sequence for enhancing expression of a target polypeptide in a particular expression system. For example, the nucleic acid can be one in which one or more codons are

altered, and often 10% or more or 20% or more of the codons are altered for optimized expression in bacteria (*e.g.*, *E. coli.*), yeast (*e.g.*, *S. cervesiae*), human (*e.g.*, 293 cells), insect, or rodent (*e.g.*, hamster) cells.

Methods for Identifying Risk of osteoarthritis

[0080] Methods for prognosing and diagnosing osteoarthritis are included herein. These methods include detecting the presence or absence of one or more polymorphic variations in a nucleotide sequence associated with osteoarthritis, such as variants in or around the loci set forth herein, or a substantially identical sequence thereof, in a sample from a subject, where the presence of a polymorphic variant described herein is indicative of a risk of osteoarthritis. Determining a risk of osteoarthritis sometimes refers to determining whether an individual is at an increased risk of osteoarthritis (*e.g.*, intermediate risk or higher risk).

[0081] Thus, featured herein is a method for identifying a subject who is at risk of osteoarthritis, which comprises detecting an aberration associated with osteoarthritis in a nucleic acid sample from the subject. An embodiment is a method for detecting a risk of osteoarthritis in a subject, which comprises detecting the presence or absence of a polymorphic variation associated with osteoarthritis at a polymorphic site in a nucleotide sequence in a nucleic acid sample from a subject, where the nucleotide sequence comprises a polynucleotide sequence selected from the group consisting of: (a) a nucleotide sequence of SEQ ID NO: 1-7; (b) a nucleotide sequence which encodes a polypeptide consisting of an amino acid sequence encoded by a nucleotide sequence of SEQ ID NO: 1-7; (c) a nucleotide sequence which encodes a polypeptide that is 90% or more identical to an amino acid sequence encoded by a nucleotide sequence of SEQ ID NO: 1-7, or a nucleotide sequence about 90% or more identical to a nucleotide sequence of SEQ ID NO: 1-7; and (d) a fragment of a nucleotide sequence of (a), (b), or (c) comprising the polymorphic site; whereby the presence of the polymorphic variation is indicative of a predisposition to osteoarthritis in the subject. In certain embodiments, polymorphic variants at the positions described herein are detected for determining a risk of osteoarthritis, and polymorphic variants at positions in linkage disequilibrium with these positions are detected for determining a risk of osteoarthritis. As used herein, "SEQ ID NO: 1-7" refers to individual sequences in SEQ ID NO: 1, 2, 3, 4, 5, 6 or 7, each sequence being separately applicable to embodiments described herein.

[0082] Risk of osteoarthritis sometimes is expressed as a probability, such as an odds ratio, percentage, or risk factor. Risk often is based upon the presence or absence of one or more polymorphic variants described herein, and also may be based in part upon phenotypic traits of the individual being tested. Methods for calculating risk based upon patient data are well known (*see, e.g.*, Agresti, *Categorical Data Analysis*, 2nd Ed. 2002. Wiley). Allelotyping and genotyping analyses may be carried out in populations other than those exemplified herein to enhance the predictive power of the prognostic

method. These further analyses are executed in view of the exemplified procedures described herein, and may be based upon the same polymorphic variations or additional polymorphic variations.

[0083] In certain embodiments, determining the presence of a combination of two or more polymorphic variants associated with osteoarthritis in one or more genetic loci (e.g., one or more genes) of the sample is determined to identify, quantify and/or estimate, risk of osteoarthritis. The risk often is the probability of having or developing osteoarthritis. The risk sometimes is expressed as a relative risk with respect to a population average risk of osteoarthritis, and sometimes is expressed as a relative risk with respect to the lowest risk group. Such relative risk assessments often are based upon penetrance values determined by statistical methods, and are particularly useful to clinicians and insurance companies for assessing risk of osteoarthritis (e.g., a clinician can target appropriate detection, prevention and therapeutic regimens to a patient after determining the patient's risk of osteoarthritis, and an insurance company can fine tune actuarial tables based upon population genotype assessments of osteoarthritis risk). Risk of osteoarthritis sometimes is expressed as an odds ratio, which is the odds of a particular person having a genotype has or will develop osteoarthritis with respect to another genotype group (e.g., the most disease protective genotype or population average). In related embodiments, the determination is utilized to identify a subject at risk of osteoarthritis. In an embodiment, two or more polymorphic variations are detected in two or more regions in human genomic DNA associated with increased risk of osteoarthritis, such as a locus containing a *APOL3*, for example. In certain embodiments, 3 or more, or 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19 or 20 or more polymorphic variants are detected in the sample. In specific embodiments, polymorphic variants are detected in a *APOL3* region, for example. In certain embodiments, polymorphic variants are detected at other genetic loci (e.g., the polymorphic variants can be detected in *APOL3* in addition to other loci or only in other loci), where the other loci include but are not limited to those described in concurrently-filed patent applications having attorney docket number 524593008700, 524593008800, 524593008900, 524593009000 or 524593009100, each of which is incorporated herein by reference in its entirety.

[0084] Results from prognostic tests may be combined with other test results to diagnose osteoarthritis. For example, prognostic results may be gathered, a patient sample may be ordered based on a determined predisposition to osteoarthritis, the patient sample is analyzed, and the results of the analysis may be utilized to diagnose osteoarthritis. Also osteoarthritis diagnostic method can be developed from studies used to generate prognostic methods in which populations are stratified into subpopulations having different progressions of osteoarthritis. In another embodiment, prognostic results may be gathered, a patient's risk factors for developing osteoarthritis (*e.g.*, age, weight, race, diet) analyzed, and a patient sample may be ordered based on a determined predisposition to osteoarthritis.

[0085] The nucleic acid sample typically is isolated from a biological sample obtained from a subject. For example, nucleic acid can be isolated from blood, saliva, sputum, urine, cell scrapings, and

biopsy tissue. The nucleic acid sample can be isolated from a biological sample using standard techniques, such as the technique described in Example 2. As used herein, the term "subject" refers primarily to humans but also refers to other mammals such as dogs, cats, and ungulates (*e.g.*, cattle, sheep, and swine). Subjects also include avians (*e.g.*, chickens and turkeys), reptiles, and fish (*e.g.*, salmon), as embodiments described herein can be adapted to nucleic acid samples isolated from any of these organisms. The nucleic acid sample may be isolated from the subject and then directly utilized in a method for determining the presence of a polymorphic variant, or alternatively, the sample may be isolated and then stored (*e.g.*, frozen) for a period of time before being subjected to analysis.

[0086] The presence or absence of a polymorphic variant is determined using one or both chromosomal complements represented in the nucleic acid sample. Determining the presence or absence of a polymorphic variant in both chromosomal complements represented in a nucleic acid sample from a subject having a copy of each chromosome is useful for determining the zygosity of an individual for the polymorphic variant (*i.e.*, whether the individual is homozygous or heterozygous for the polymorphic variant). Any oligonucleotide-based diagnostic may be utilized to determine whether a sample includes the presence or absence of a polymorphic variant in a sample. For example, primer extension methods, ligase sequence determination methods (*e.g.*, U.S. Pat. Nos. 5,679,524 and 5,952,174, and WO 01/27326), mismatch sequence determination methods (*e.g.*, U.S. Pat. Nos. 5,851,770; 5,958,692; 6,110,684; and 6,183,958), microarray sequence determination methods, restriction fragment length polymorphism (RFLP), single strand conformation polymorphism detection (SSCP) (*e.g.*, U.S. Pat. Nos. 5,891,625 and 6,013,499), PCR-based assays (*e.g.*, TAQMAN® PCR System (Applied Biosystems)), and nucleotide sequencing methods may be used.

[0087] Oligonucleotide extension methods typically involve providing a pair of oligonucleotide primers in a polymerase chain reaction (PCR) or in other nucleic acid amplification methods for the purpose of amplifying a region from the nucleic acid sample that comprises the polymorphic variation. One oligonucleotide primer is complementary to a region 3' of the polymorphism and the other is complementary to a region 5' of the polymorphism. A PCR primer pair may be used in methods disclosed in U.S. Pat. Nos. 4,683,195; 4,683,202, 4,965,188; 5,656,493; 5,998,143; 6,140,054; WO 01/27327; and WO 01/27329 for example. PCR primer pairs may also be used in any commercially available machines that perform PCR, such as any of the GENEAMP® Systems available from Applied Biosystems. Also, those of ordinary skill in the art will be able to design oligonucleotide primers based upon a *APOL3* nucleotide sequence using knowledge available in the art.

[0088] Also provided is an extension oligonucleotide that hybridizes to the amplified fragment adjacent to the polymorphic variation. As used herein, the term "adjacent" refers to the 3' end of the extension oligonucleotide being often 1 nucleotide from the 5' end of the polymorphic site, and sometimes 2, 3, 4, 5, 6, 7, 8, 9, or 10 nucleotides from the 5' end of the polymorphic site, in the nucleic

acid when the extension oligonucleotide is hybridized to the nucleic acid. The extension oligonucleotide then is extended by one or more nucleotides, and the number and/or type of nucleotides that are added to the extension oligonucleotide determine whether the polymorphic variant is present. Oligonucleotide extension methods are disclosed, for example, in U.S. Pat. Nos. 4,656,127; 4,851,331; 5,679,524; 5,834,189; 5,876,934; 5,908,755; 5,912,118; 5,976,802; 5,981,186; 6,004,744; 6,013,431; 6,017,702; 6,046,005; 6,087,095; 6,210,891; and WO 01/20039. Oligonucleotide extension methods using mass spectrometry are described, for example, in U.S. Pat. Nos. 5,547,835; 5,605,798; 5,691,141; 5,849,542; 5,869,242; 5,928,906; 6,043,031; and 6,194,144, and a method often utilized is described herein in Example 2.

[0089] A microarray can be utilized for determining whether a polymorphic variant is present or absent in a nucleic acid sample. A microarray may include any oligonucleotides described herein, and methods for making and using oligonucleotide microarrays suitable for diagnostic use are disclosed in U.S. Pat. Nos. 5,492,806; 5,525,464; 5,589,330; 5,695,940; 5,849,483; 6,018,041; 6,045,996; 6,136,541; 6,142,681; 6,156,501; 6,197,506; 6,223,127; 6,225,625; 6,229,911; 6,239,273; WO 00/52625; WO 01/25485; and WO 01/29259. The microarray typically comprises a solid support and the oligonucleotides may be linked to this solid support by covalent bonds or by non-covalent interactions. The oligonucleotides may also be linked to the solid support directly or by a spacer molecule. A microarray may comprise one or more oligonucleotides complementary to a polymorphic site set forth herein.

[0090] A kit also may be utilized for determining whether a polymorphic variant is present or absent in a nucleic acid sample. A kit often comprises one or more pairs of oligonucleotide primers useful for amplifying a fragment of a nucleotide sequence of SEQ ID NO: 1-7 or a substantially identical sequence thereof, where the fragment includes a polymorphic site. The kit sometimes comprises a polymerizing agent, for example, a thermostable nucleic acid polymerase such as one disclosed in U.S. Pat. Nos. 4,889,818 or 6,077,664. Also, the kit often comprises an elongation oligonucleotide that hybridizes to a *APOL3* nucleotide sequence in a nucleic acid sample adjacent to the polymorphic site. Where the kit includes an elongation oligonucleotide, it also often comprises chain elongating nucleotides, such as dATP, dTTP, dGTP, dCTP, and dITP, including analogs of dATP, dTTP, dGTP, dCTP and dITP, provided that such analogs are substrates for a thermostable nucleic acid polymerase and can be incorporated into a nucleic acid chain elongated from the extension oligonucleotide. Along with chain elongating nucleotides would be one or more chain terminating nucleotides such as ddATP, ddTTP, ddGTP, ddCTP, and the like. In an embodiment, the kit comprises one or more oligonucleotide primer pairs, a polymerizing agent, chain elongating nucleotides, at least one elongation oligonucleotide, and one or more chain terminating nucleotides. Kits optionally include buffers, vials, microtiter plates, and instructions for use.

[0091] An individual identified as being at risk of osteoarthritis may be heterozygous or homozygous with respect to the allele associated with a higher risk of osteoarthritis. A subject homozygous for an allele associated with an increased risk of osteoarthritis is at a comparatively high risk of osteoarthritis, a subject heterozygous for an allele associated with an increased risk of osteoarthritis is at a comparatively intermediate risk of osteoarthritis, and a subject homozygous for an allele associated with a decreased risk of osteoarthritis is at a comparatively low risk of osteoarthritis. A genotype may be assessed for a complementary strand, such that the complementary nucleotide at a particular position is detected.

[0092] Also featured are methods for determining risk of osteoarthritis and/or identifying a subject at risk of osteoarthritis by contacting a polypeptide or protein encoded by a *APOL3* nucleotide sequence from a subject with an antibody that specifically binds to an epitope associated with increased risk of osteoarthritis in the polypeptide.

Applications of Prognostic and Diagnostic Results to Pharmacogenomic Methods

[0093] Pharmacogenomics is a discipline that involves tailoring a treatment for a subject according to the subject's genotype as a particular treatment regimen may exert a differential effect depending upon the subject's genotype. For example, based upon the outcome of a prognostic test described herein, a clinician or physician may target pertinent information and preventative or therapeutic treatments to a subject who would be benefited by the information or treatment and avoid directing such information and treatments to a subject who would not be benefited (*e.g.*, the treatment has no therapeutic effect and/or the subject experiences adverse side effects).

[0094] The following is an example of a pharmacogenomic embodiment. A particular treatment regimen can exert a differential effect depending upon the subject's genotype. Where a candidate therapeutic exhibits a significant interaction with a major allele and a comparatively weak interaction with a minor allele (*e.g.*, an order of magnitude or greater difference in the interaction), such a therapeutic typically would not be administered to a subject genotyped as being homozygous for the minor allele, and sometimes not administered to a subject genotyped as being heterozygous for the minor allele. In another example, where a candidate therapeutic is not significantly toxic when administered to subjects who are homozygous for a major allele but is comparatively toxic when administered to subjects heterozygous or homozygous for a minor allele, the candidate therapeutic is not typically administered to subjects who are genotyped as being heterozygous or homozygous with respect to the minor allele.

[0095] The methods described herein are applicable to pharmacogenomic methods for preventing, alleviating or treating osteoarthritis. For example, a nucleic acid sample from an individual may be subjected to a prognostic test described herein. Where one or more polymorphic variations associated

with increased risk of osteoarthritis are identified in a subject, information for preventing or treating osteoarthritis and/or one or more osteoarthritis treatment regimens then may be prescribed to that subject.

[0096] In certain embodiments, a treatment or preventative regimen is specifically prescribed and/or administered to individuals who will most benefit from it based upon their risk of developing osteoarthritis assessed by the methods described herein. Thus, provided are methods for identifying a subject predisposed to osteoarthritis and then prescribing a therapeutic or preventative regimen to individuals identified as having a predisposition. Thus, certain embodiments are directed to a method for reducing osteoarthritis in a subject, which comprises: detecting the presence or absence of a polymorphic variant associated with osteoarthritis in a nucleotide sequence in a nucleic acid sample from a subject, where the nucleotide sequence comprises a polynucleotide sequence selected from the group consisting of: (a) a nucleotide sequence of SEQ ID NO: 1-7; (b) a nucleotide sequence which encodes a polypeptide consisting of an amino acid sequence encoded by a nucleotide sequence of SEQ ID NO: 1-7; (c) a nucleotide sequence which encodes a polypeptide that is 90% or more identical to an amino acid sequence encoded by a nucleotide sequence of SEQ ID NO: 1-7, or a nucleotide sequence about 90% or more identical to a nucleotide sequence of SEQ ID NO: 1-7; and (d) a fragment of a polynucleotide sequence of (a), (b), or (c); and prescribing or administering a treatment regimen to a subject from whom the sample originated where the presence of a polymorphic variation associated with osteoarthritis is detected in the nucleotide sequence. In these methods, predisposition results may be utilized in combination with other test results to diagnose osteoarthritis.

[0097] Certain preventative treatments often are prescribed to subjects having a predisposition to osteoarthritis and where the subject is diagnosed with osteoarthritis or is diagnosed as having symptoms indicative of an early stage of osteoarthritis. The treatment sometimes is preventative (*e.g.*, is prescribed or administered to reduce the probability that osteoarthritis arises or progresses), sometimes is therapeutic, and sometimes delays, alleviates or halts the progression of osteoarthritis. Any known preventative or therapeutic treatment for alleviating or preventing the occurrence of osteoarthritis is prescribed and/or administered. For example, the treatment often is directed to decreasing pain and improving joint movement. Examples of OA treatments include exercises to keep joints flexible and improve muscle strength. Different medications to control pain, including corticosteroids and nonsteroidal anti-inflammatory drugs (NSAIDs, e.g., Voltaren); cyclooxygenase-2 (COX-2) inhibitors (e.g., Celebrex, Vioxx, Mobic, and Bextra); monoclonal antibodies (e.g., Remicade); tumor necrosis factor inhibitors (e.g., Enbrel); or injections of glucocorticoids, hyaluronic acid or chondrotin sulfate into joints that are inflamed and not responsive to NSAIDS. Orally administered chondroitin sulfate also may be used as a therapeutic, as it may increase hyaluronic acid levels and viscosity of synovial fluid, and decrease collagenase levels in synovial fluid. Also, glucosamine can serve as an OA therapeutic as delivering it into joints may inhibit enzymes involved in cartilage degradation and enhance the

production of hyaluronic acid. For mild pain without inflammation, acetaminophen may be used. Other treatments include: heat/cold therapy for temporary pain relief; joint protection to prevent strain or stress on painful joints; surgery to relieve chronic pain in damaged joints; and weight control to prevent extra stress on weight-bearing joints.

[0098] As therapeutic approaches for treating osteoarthritis continue to evolve and improve, the goal of treatments for osteoarthritis related disorders is to intervene even before clinical signs first manifest. Thus, genetic markers associated with susceptibility to osteoarthritis prove useful for early diagnosis, prevention and treatment of osteoarthritis.

[0099] As osteoarthritis preventative and treatment information can be specifically targeted to subjects in need thereof (*e.g.*, those at risk of developing osteoarthritis or those in an early stage of osteoarthritis), provided herein is a method for preventing or reducing the risk of developing osteoarthritis in a subject, which comprises: (a) detecting the presence or absence of a polymorphic variation associated with osteoarthritis at a polymorphic site in a nucleotide sequence in a nucleic acid sample from a subject; (b) identifying a subject with a predisposition to osteoarthritis, whereby the presence of the polymorphic variation is indicative of a predisposition to osteoarthritis in the subject; and (c) if such a predisposition is identified, providing the subject with information about methods or products to prevent or reduce osteoarthritis or to delay the onset of osteoarthritis. Also provided is a method of targeting information or advertising to a subpopulation of a human population based on the subpopulation being genetically predisposed to a disease or condition, which comprises: (a) detecting the presence or absence of a polymorphic variation associated with osteoarthritis at a polymorphic site in a nucleotide sequence in a nucleic acid sample from a subject; (b) identifying the subpopulation of subjects in which the polymorphic variation is associated with osteoarthritis; and (c) providing information only to the subpopulation of subjects about a particular product which may be obtained and consumed or applied by the subject to help prevent or delay onset of the disease or condition.

[0100] Pharmacogenomics methods also may be used to analyze and predict a response to osteoarthritis treatment or a drug. For example, if pharmacogenomics analysis indicates a likelihood that an individual will respond positively to osteoarthritis treatment with a particular drug, the drug may be administered to the individual. Conversely, if the analysis indicates that an individual is likely to respond negatively to treatment with a particular drug, an alternative course of treatment may be prescribed. A negative response may be defined as either the absence of an efficacious response or the presence of toxic side effects. The response to a therapeutic treatment can be predicted in a background study in which subjects in any of the following populations are genotyped: a population that responds favorably to a treatment regimen, a population that does not respond significantly to a treatment regimen, and a population that responds adversely to a treatment regimen (*e.g.*, exhibits one or more side effects). These populations are provided as examples and other populations and subpopulations may be analyzed. Based

upon the results of these analyses, a subject is genotyped to predict whether he or she will respond favorably to a treatment regimen, not respond significantly to a treatment regimen, or respond adversely to a treatment regimen.

[0101] The tests described herein also are applicable to clinical drug trials. One or more polymorphic variants indicative of response to an agent for treating osteoarthritis or to side effects to an agent for treating osteoarthritis may be identified using the methods described herein. Thereafter, potential participants in clinical trials of such an agent may be screened to identify those individuals most likely to respond favorably to the drug and exclude those likely to experience side effects. In that way, the effectiveness of drug treatment may be measured in individuals who respond positively to the drug, without lowering the measurement as a result of the inclusion of individuals who are unlikely to respond positively in the study and without risking undesirable safety problems.

[0102] Thus, another embodiment is a method of selecting an individual for inclusion in a clinical trial of a treatment or drug comprising the steps of: (a) obtaining a nucleic acid sample from an individual; (b) determining the identity of a polymorphic variation which is associated with a positive response to the treatment or the drug, or at least one polymorphic variation which is associated with a negative response to the treatment or the drug in the nucleic acid sample, and (c) including the individual in the clinical trial if the nucleic acid sample contains said polymorphic variation associated with a positive response to the treatment or the drug or if the nucleic acid sample lacks said polymorphic variation associated with a negative response to the treatment or the drug. In addition, the methods described herein for selecting an individual for inclusion in a clinical trial of a treatment or drug encompass methods with any further limitation described in this disclosure, or those following, specified alone or in any combination. The polymorphic variation may be in a sequence selected individually or in any combination from the group consisting of (i) a nucleotide sequence of SEQ ID NO: 1-7; (ii) a nucleotide sequence which encodes a polypeptide consisting of an amino acid sequence encoded by a nucleotide sequence of SEQ ID NO: 1-7; (iii) a nucleotide sequence which encodes a polypeptide that is 90% or more identical to an amino acid sequence encoded by a nucleotide sequence of SEQ ID NO: 1-7, or a nucleotide sequence about 90% or more identical to a nucleotide sequence of SEQ ID NO: 1-7; and (iv) a fragment of a polynucleotide sequence of (i), (ii), or (iii) comprising the polymorphic site. The including step (c) optionally comprises administering the drug or the treatment to the individual if the nucleic acid sample contains the polymorphic variation associated with a positive response to the treatment or the drug and the nucleic acid sample lacks said biallelic marker associated with a negative response to the treatment or the drug.

[0103] Also provided herein is a method of partnering between a diagnostic/prognostic testing provider and a provider of a consumable product, which comprises: (a) the diagnostic/prognostic testing provider detects the presence or absence of a polymorphic variation associated with osteoarthritis at a

polymorphic site in a nucleotide sequence in a nucleic acid sample from a subject; (b) the diagnostic/prognostic testing provider identifies the subpopulation of subjects in which the polymorphic variation is associated with osteoarthritis; (c) the diagnostic/prognostic testing provider forwards information to the subpopulation of subjects about a particular product which may be obtained and consumed or applied by the subject to help prevent or delay onset of the disease or condition; and (d) the provider of a consumable product forwards to the diagnostic test provider a fee every time the diagnostic/prognostic test provider forwards information to the subject as set forth in step (c) above.

Compositions Comprising Osteoarthritis-Directed Molecules

[0104] Featured herein is a composition comprising a cell from a subject having osteoarthritis or at risk of osteoarthritis and one or more molecules specifically directed and targeted to a nucleic acid comprising a *APOL3* nucleotide sequence or amino acid sequence. Such directed molecules include, but are not limited to, a compound that binds to a *APOL3* nucleotide sequence or amino acid sequence referenced herein; a RNAi or siRNA molecule having a strand complementary or substantially complementary to a *APOL3* nucleotide sequence (e.g., hybridizes to a *APOL3* nucleotide sequence under conditions of high stringency); an antisense nucleic acid complementary or substantially complementary to an RNA encoded by a *APOL3* nucleotide sequence (e.g., hybridizes to a *APOL3* nucleotide sequence under conditions of high stringency); a ribozyme that hybridizes to a *APOL3* nucleotide sequence (e.g., hybridizes to a *APOL3* nucleotide sequence under conditions of high stringency); a nucleic acid aptamer that specifically binds a polypeptide encoded by *APOL3* nucleotide sequence; and an antibody that specifically binds to a polypeptide encoded by *APOL3* nucleotide sequence or binds to a nucleic acid having such a nucleotide sequence. In specific embodiments, the osteoarthritis directed molecule interacts with a nucleic acid or polypeptide variant associated with osteoarthritis, such as variants referenced herein. In other embodiments, the osteoarthritis directed molecule interacts with a polypeptide involved in a signal pathway of a polypeptide encoded by a *APOL3* nucleotide sequence, or a nucleic acid comprising such a nucleotide sequence.

[0105] Compositions sometimes include an adjuvant known to stimulate an immune response, and in certain embodiments, an adjuvant that stimulates a T-cell lymphocyte response. Adjuvants are known, including but not limited to an aluminum adjuvant (e.g., aluminum hydroxide); a cytokine adjuvant or adjuvant that stimulates a cytokine response (e.g., interleukin (IL)-12 and/or gamma-interferon cytokines); a Freund-type mineral oil adjuvant emulsion (e.g., Freund's complete or incomplete adjuvant); a synthetic lipoid compound; a copolymer adjuvant (e.g., TitreMax); a saponin; Quil A; a liposome; an oil-in-water emulsion (e.g., an emulsion stabilized by Tween 80 and pluronic polyoxyethlene/polyoxypropylene block copolymer (Syntex Adjuvant Formulation); TitreMax; detoxified endotoxin (MPL) and mycobacterial cell wall components (TDW, CWS) in 2% squalene (Ribi

Adjuvant System)); a muramyl dipeptide; an immune-stimulating complex (ISCOM, e.g., an Ag-modified saponin/cholesterol micelle that forms stable cage-like structure); an aqueous phase adjuvant that does not have a depot effect (e.g., Gerbu adjuvant); a carbohydrate polymer (e.g., AdjuPrime); L-tyrosine; a manide-oleate compound (e.g., Montanide); an ethylene-vinyl acetate copolymer (e.g., Elvax 40W1,2); or lipid A, for example. Such compositions are useful for generating an immune response against osteoarthritis directed molecule (e.g., an HLA-binding subsequence within a polypeptide encoded by a *APOL3* nucleotide sequence). In such methods, a peptide having an amino acid subsequence of a polypeptide encoded by a *APOL3* nucleotide sequence is delivered to a subject, where the subsequence binds to an HLA molecule and induces a CTL lymphocyte response. The peptide sometimes is delivered to the subject as an isolated peptide or as a minigene in a plasmid that encodes the peptide. Methods for identifying HLA-binding subsequences in such polypeptides are known (see e.g., publication WO02/20616 and PCT application US98/01373 for methods of identifying such sequences).

[0106] The cell may be in a group of cells cultured *in vitro* or in a tissue maintained *in vitro* or present in an animal *in vivo* (e.g., a rat, mouse, ape or human). In certain embodiments, a composition comprises a component from a cell such as a nucleic acid molecule (e.g., genomic DNA), a protein mixture or isolated protein, for example. The aforementioned compositions have utility in diagnostic, prognostic and pharmacogenomic methods described previously and in therapeutics described hereafter. Certain osteoarthritis directed molecules are described in greater detail below.

<u>Compounds</u>

[0107] Compounds can be obtained using any of the numerous approaches in combinatorial library methods known in the art, including: biological libraries; peptoid libraries (libraries of molecules having the functionalities of peptides, but with a novel, non-peptide backbone which are resistant to enzymatic degradation but which nevertheless remain bioactive (see, e.g., Zuckermann et al., J. Med. Chem.37: 2678-85 (1994)); spatially addressable parallel solid phase or solution phase libraries; synthetic library methods requiring deconvolution; "one-bead one-compound" library methods; and synthetic library methods using affinity chromatography selection. Biological library and peptoid library approaches are typically limited to peptide libraries, while the other approaches are applicable to peptide, non-peptide oligomer or small molecule libraries of compounds (Lam, Anticancer Drug Des. 12: 145, (1997)). Examples of methods for synthesizing molecular libraries are described, for example, in DeWitt et al., Proc. Natl. Acad. Sci. U.S.A. 90: 6909 (1993); Erb et al., Proc. Natl. Acad. Sci. USA 91: 11422 (1994); Zuckermann et al., J. Med. Chem. 37: 2678 (1994); Cho et al., Science 261: 1303 (1993); Carrell et al., Angew. Chem. Int. Ed. Engl. 33: 2059 (1994); Carell et al., Angew. Chem. Int. Ed. Engl. 33: 2061 (1994); and in Gallop et al., J. Med. Chem. 37: 1233 (1994).

[0108] Libraries of compounds may be presented in solution (e.g., Houghten, Biotechniques 13: 412-421 (1992)), or on beads (Lam, Nature 354: 82-84 (1991)), chips (Fodor, Nature 364: 555-556 (1993)), bacteria or spores (Ladner, United States Patent No. 5,223,409), plasmids (Cull et al., Proc. Natl. Acad. Sci. USA 89: 1865-1869 (1992)) or on phage (Scott and Smith, Science 249: 386-390 (1990); Devlin, Science 249: 404-406 (1990); Cwirla et al., Proc. Natl. Acad. Sci. 87: 6378-6382 (1990); Felici, J. Mol. Biol. 222: 301-310 (1991); Ladner supra.).

[0109] A compound sometimes alters expression and sometimes alters activity of a polypeptide target and may be a small molecule. Small molecules include, but are not limited to, peptides, peptidomimetics (e.g., peptoids), amino acids, amino acid analogs, polynucleotides, polynucleotide analogs, nucleotides, nucleotide analogs, organic or inorganic compounds (i.e., including heteroorganic and organometallic compounds) having a molecular weight less than about 10,000 grams per mole, organic or inorganic compounds having a molecular weight less than about 5,000 grams per mole, organic or inorganic compounds having a molecular weight less than about 1,000 grams per mole, organic or inorganic compounds having a molecular weight less than about 500 grams per mole, and salts, esters, and other pharmaceutically acceptable forms of such compounds.

<u>Antisense Nucleic Acid Molecules, Ribozymes, RNAi, siRNA and Modified Nucleic Acid Molecules</u>

[0110] An "antisense" nucleic acid refers to a nucleotide sequence complementary to a "sense" nucleic acid encoding a polypeptide, e.g., complementary to the coding strand of a double-stranded cDNA molecule or complementary to an mRNA sequence. The antisense nucleic acid can be complementary to an entire coding strand, or to a portion thereof or a substantially identical sequence thereof. In another embodiment, the antisense nucleic acid molecule is antisense to a "noncoding region" of the coding strand of a nucleotide sequence (e.g., 5' and 3' untranslated regions in SEQ ID NO: 1).

[0111] An antisense nucleic acid can be designed such that it is complementary to the entire coding region of an mRNA encoded by a nucleotide sequence (e.g., SEQ ID NO: 1), and often the antisense nucleic acid is an oligonucleotide antisense to only a portion of a coding or noncoding region of the mRNA. For example, the antisense oligonucleotide can be complementary to the region surrounding the translation start site of the mRNA, e.g., between the -10 and +10 regions of the target gene nucleotide sequence of interest. An antisense oligonucleotide can be, for example, about 7, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, or more nucleotides in length. The antisense nucleic acids, which include the ribozymes described hereafter, can be designed to target a *APOL3* nucleotide sequence, often a variant associated with osteoarthritis, or a substantially identical sequence thereof. Among the variants, minor alleles and major alleles can be targeted, and those associated with a higher risk of osteoarthritis are often designed, tested, and administered to subjects.

[0112] An antisense nucleic acid can be constructed using chemical synthesis and enzymatic ligation reactions using standard procedures. For example, an antisense nucleic acid (e.g., an antisense oligonucleotide) can be chemically synthesized using naturally occurring nucleotides or variously modified nucleotides designed to increase the biological stability of the molecules or to increase the physical stability of the duplex formed between the antisense and sense nucleic acids, e.g., phosphorothioate derivatives and acridine substituted nucleotides can be used. Antisense nucleic acid also can be produced biologically using an expression vector into which a nucleic acid has been subcloned in an antisense orientation (i.e., RNA transcribed from the inserted nucleic acid will be of an antisense orientation to a target nucleic acid of interest, described further in the following subsection).

[0113] When utilized as therapeutics, antisense nucleic acids typically are administered to a subject (e.g., by direct injection at a tissue site) or generated in situ such that they hybridize with or bind to cellular mRNA and/or genomic DNA encoding a polypeptide and thereby inhibit expression of the polypeptide, for example, by inhibiting transcription and/or translation. Alternatively, antisense nucleic acid molecules can be modified to target selected cells and then are administered systemically. For systemic administration, antisense molecules can be modified such that they specifically bind to receptors or antigens expressed on a selected cell surface, for example, by linking antisense nucleic acid molecules to peptides or antibodies which bind to cell surface receptors or antigens. Antisense nucleic acid molecules can also be delivered to cells using the vectors described herein. Sufficient intracellular concentrations of antisense molecules are achieved by incorporating a strong promoter, such as a pol II or pol III promoter, in the vector construct.

[0114] Antisense nucleic acid molecules sometimes are alpha-anomeric nucleic acid molecules. An alpha-anomeric nucleic acid molecule forms specific double-stranded hybrids with complementary RNA in which, contrary to the usual beta-units, the strands run parallel to each other (Gaultier et al., Nucleic Acids. Res. 15: 6625-6641 (1987)). Antisense nucleic acid molecules can also comprise a 2'-o-methylribonucleotide (Inoue et al., Nucleic Acids Res. 15: 6131-6148 (1987)) or a chimeric RNA-DNA analogue (Inoue et al., FEBS Lett. 215: 327-330 (1987)). Antisense nucleic acids sometimes are composed of DNA or PNA or any other nucleic acid derivatives described previously.

[0115] In another embodiment, an antisense nucleic acid is a ribozyme. A ribozyme having specificity for a *APOL3* nucleotide sequence can include one or more sequences complementary to such a nucleotide sequence, and a sequence having a known catalytic region responsible for mRNA cleavage (see e.g., U.S. Pat. No. 5,093,246 or Haselhoff and Gerlach, Nature 334: 585-591 (1988)). For example, a derivative of a Tetrahymena L-19 IVS RNA is sometimes utilized in which the nucleotide sequence of the active site is complementary to the nucleotide sequence to be cleaved in a mRNA (see e.g., Cech et al. U.S. Patent No. 4,987,071; and Cech et al. U.S. Patent No. 5,116,742). Also, target mRNA sequences

can be used to select a catalytic RNA having a specific ribonuclease activity from a pool of RNA molecules (see e.g., Bartel & Szostak, Science 261: 1411-1418 (1993)).

[0116] Osteoarthritis directed molecules include in certain embodiments nucleic acids that can form triple helix structures with a *APOL3* nucleotide sequence, or a substantially identical sequence thereof, especially one that includes a regulatory region that controls expression of a polypeptide. Gene expression can be inhibited by targeting nucleotide sequences complementary to the regulatory region of a nucleotide sequence referenced herein or a substantially identical sequence (e.g., promoter and/or enhancers) to form triple helical structures that prevent transcription of a gene in target cells (see e.g., Helene, Anticancer Drug Des. 6(6): 569-84 (1991); Helene et al., Ann. N.Y. Acad. Sci. 660: 27-36 (1992); and Maher, Bioassays 14(12): 807-15 (1992). Potential sequences that can be targeted for triple helix formation can be increased by creating a so-called "switchback" nucleic acid molecule. Switchback molecules are synthesized in an alternating 5'-3', 3'-5' manner, such that they base pair with first one strand of a duplex and then the other, eliminating the necessity for a sizeable stretch of either purines or pyrimidines to be present on one strand of a duplex.

[0117] Osteoarthritis directed molecules include RNAi and siRNA nucleic acids. Gene expression may be inhibited by the introduction of double-stranded RNA (dsRNA), which induces potent and specific gene silencing, a phenomenon called RNA interference or RNAi. See, e.g., Fire et al., US Patent Number 6,506,559; Tuschl et al. PCT International Publication No. WO 01/75164; Kay et al. PCT International Publication No. WO 03/010180A1; or Bosher JM, Labouesse, Nat Cell Biol 2000 Feb;2(2):E31-6. This process has been improved by decreasing the size of the double-stranded RNA to 20-24 base pairs (to create small-interfering RNAs or siRNAs) that "switched off" genes in mammalian cells without initiating an acute phase response, i.e., a host defense mechanism that often results in cell death (see, e.g., Caplen et al. Proc Natl Acad Sci U S A. 2001 Aug 14;98(17):9742-7 and Elbashir et al. Methods 2002 Feb;26(2):199-213). There is increasing evidence of post-transcriptional gene silencing by RNA interference (RNAi) for inhibiting targeted expression in mammalian cells at the mRNA level, in human cells. There is additional evidence of effective methods for inhibiting the proliferation and migration of tumor cells in human patients, and for inhibiting metastatic cancer development (see, e.g., U.S. Patent Application No. US2001000993183; Caplen et al. Proc Natl Acad Sci U S A; and Abderrahmani et al. Mol Cell Biol 2001 Nov21(21):7256-67).

[0118] An "siRNA" or "RNAi" refers to a nucleic acid that forms a double stranded RNA and has the ability to reduce or inhibit expression of a gene or target gene when the siRNA is delivered to or expressed in the same cell as the gene or target gene. "siRNA" refers to short double-stranded RNA formed by the complementary strands. Complementary portions of the siRNA that hybridize to form the double stranded molecule often have substantial or complete identity to the target molecule sequence. In

one embodiment, an siRNA refers to a nucleic acid that has substantial or complete identity to a target gene and forms a double stranded siRNA.

[0119] When designing the siRNA molecules, the targeted region often is selected from a given DNA sequence beginning 50 to 100 nucleotides downstream of the start codon. See, e.g., Elbashir et al,. Methods 26:199-213 (2002). Initially, 5' or 3' UTRs and regions nearby the start codon were avoided assuming that UTR-binding proteins and/or translation initiation complexes may interfere with binding of the siRNP or RISC endonuclease complex. Sometimes regions of the target 23 nucleotides in length conforming to the sequence motif AA(N19)TT (N, an nucleotide), and regions with approximately 30% to 70% G/C-content (often about 50% G/C-content) often are selected. If no suitable sequences are found, the search often is extended using the motif NA(N21). The sequence of the sense siRNA sometimes corresponds to (N19) TT or N21 (position 3 to 23 of the 23-nt motif), respectively. In the latter case, the 3' end of the sense siRNA often is converted to TT. The rationale for this sequence conversion is to generate a symmetric duplex with respect to the sequence composition of the sense and antisense 3' overhangs. The antisense siRNA is synthesized as the complement to position 1 to 21 of the 23-nt motif. Because position 1 of the 23-nt motif is not recognized sequence-specifically by the antisense siRNA, the 3'-most nucleotide residue of the antisense siRNA can be chosen deliberately. However, the penultimate nucleotide of the antisense siRNA (complementary to position 2 of the 23-nt motif) often is complementary to the targeted sequence. For simplifying chemical synthesis, TT often is utilized. siRNAs corresponding to the target motif NAR(N17)YNN, where R is purine (A,G) and Y is pyrimidine (C,U), often are selected. Respective 21 nucleotide sense and antisense siRNAs often begin with a purine nucleotide and can also be expressed from pol III expression vectors without a change in targeting site. Expression of RNAs from pol III promoters often is efficient when the first transcribed nucleotide is a purine.

[0120] The sequence of the siRNA can correspond to the full length target gene, or a subsequence thereof. Often, the siRNA is about 15 to about 50 nucleotides in length (e.g., each complementary sequence of the double stranded siRNA is 15-50 nucleotides in length, and the double stranded siRNA is about 15-50 base pairs in length, sometimes about 20-30 nucleotides in length or about 20-25 nucleotides in length, e.g., 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, or 30 nucleotides in length. The siRNA sometimes is about 21 nucleotides in length. Methods of using siRNA are well known in the art, and specific siRNA molecules may be purchased from a number of companies including Dharmacon Research, Inc.

[0121] Antisense, ribozyme, RNAi and siRNA nucleic acids can be altered to form modified nucleic acid molecules. The nucleic acids can be altered at base moieties, sugar moieties or phosphate backbone moieties to improve stability, hybridization, or solubility of the molecule. For example, the deoxyribose phosphate backbone of nucleic acid molecules can be modified to generate peptide nucleic acids (see Hyrup et al., Bioorganic & Medicinal Chemistry 4 (1): 5-23 (1996)). As used herein, the terms "peptide

nucleic acid" or "PNA" refers to a nucleic acid mimic such as a DNA mimic, in which the deoxyribose phosphate backbone is replaced by a pseudopeptide backbone and only the four natural nucleobases are retained. The neutral backbone of a PNA can allow for specific hybridization to DNA and RNA under conditions of low ionic strength. Synthesis of PNA oligomers can be performed using standard solid phase peptide synthesis protocols as described, for example, in Hyrup et al., (1996) supra and Perry-O'Keefe et al., Proc. Natl. Acad. Sci. 93: 14670-675 (1996).

[0122] PNA nucleic acids can be used in prognostic, diagnostic, and therapeutic applications. For example, PNAs can be used as antisense or antigene agents for sequence-specific modulation of gene expression by, for example, inducing transcription or translation arrest or inhibiting replication. PNA nucleic acid molecules can also be used in the analysis of single base pair mutations in a gene, (e.g., by PNA-directed PCR clamping); as "artificial restriction enzymes" when used in combination with other enzymes, (e.g., S1 nucleases (Hyrup (1996) supra)); or as probes or primers for DNA sequencing or hybridization (Hyrup et al., (1996) supra; Perry-O'Keefe supra).

[0123] In other embodiments, oligonucleotides may include other appended groups such as peptides (e.g., for targeting host cell receptors in vivo), or agents facilitating transport across cell membranes (see e.g., Letsinger et al., Proc. Natl. Acad. Sci. USA 86: 6553-6556 (1989); Lemaitre et al., Proc. Natl. Acad. Sci. USA 84: 648-652 (1987); PCT Publication No. W088/09810) or the blood-brain barrier (see, e.g., PCT Publication No. W089/10134). In addition, oligonucleotides can be modified with hybridization-triggered cleavage agents (See, e.g., Krol et al., Bio-Techniques 6: 958-976 (1988)) or intercalating agents. (See, e.g., Zon, Pharm. Res. 5: 539-549 (1988)). To this end, the oligonucleotide may be conjugated to another molecule, (e.g., a peptide, hybridization triggered cross-linking agent, transport agent, or hybridization-triggered cleavage agent).

[0124] Also included herein are molecular beacon oligonucleotide primer and probe molecules having one or more regions complementary to a *APOL3* nucleotide sequence, or a substantially identical sequence thereof, two complementary regions one having a fluorophore and one a quencher such that the molecular beacon is useful for quantifying the presence of the nucleic acid in a sample. Molecular beacon nucleic acids are described, for example, in Lizardi et al., U.S. Patent No. 5,854,033; Nazarenko et al., U.S. Patent No. 5,866,336, and Livak et al., U.S. Patent 5,876,930.

<u>Antibodies</u>

[0125] The term "antibody" as used herein refers to an immunoglobulin molecule or immunologically active portion thereof, i.e., an antigen-binding portion. Examples of immunologically active portions of immunoglobulin molecules include F(ab) and $F(ab')_2$ fragments which can be generated by treating the antibody with an enzyme such as pepsin. An antibody sometimes is a polyclonal, monoclonal, recombinant (e.g., a chimeric or humanized), fully human, non-human (e.g.,

murine), or a single chain antibody. An antibody may have effector function and can fix complement, and is sometimes coupled to a toxin or imaging agent.

[0126] A full-length polypeptide or antigenic peptide fragment encoded by a nucleotide sequence referenced herein can be used as an immunogen or can be used to identify antibodies made with other immunogens, e.g., cells, membrane preparations, and the like. An antigenic peptide often includes at least 8 amino acid residues of the amino acid sequences encoded by a nucleotide sequence referenced herein, or substantially identical sequence thereof, and encompasses an epitope. Antigenic peptides sometimes include 10 or more amino acids, 15 or more amino acids, 20 or more amino acids, or 30 or more amino acids. Hydrophilic and hydrophobic fragments of polypeptides sometimes are used as immunogens.

[0127] Epitopes encompassed by the antigenic peptide are regions located on the surface of the polypeptide (e.g., hydrophilic regions) as well as regions with high antigenicity. For example, an Emini surface probability analysis of the human polypeptide sequence can be used to indicate the regions that have a particularly high probability of being localized to the surface of the polypeptide and are thus likely to constitute surface residues useful for targeting antibody production. The antibody may bind an epitope on any domain or region on polypeptides described herein.

[0128] Also, chimeric, humanized, and completely human antibodies are useful for applications which include repeated administration to subjects. Chimeric and humanized monoclonal antibodies, comprising both human and non-human portions, can be made using standard recombinant DNA techniques. Such chimeric and humanized monoclonal antibodies can be produced by recombinant DNA techniques known in the art, for example using methods described in Robinson et al International Application No. PCT/US86/02269; Akira, et al European Patent Application 184,187; Taniguchi, M., European Patent Application 171,496; Morrison et al European Patent Application 173,494; Neuberger et al PCT International Publication No. WO 86/01533; Cabilly et al U.S. Patent No. 4,816,567; Cabilly et al European Patent Application 125,023; Better et al., Science 240: 1041-1043 (1988); Liu et al., Proc. Natl. Acad. Sci. USA 84: 3439-3443 (1987); Liu et al., J. Immunol. 139: 3521-3526 (1987); Sun et al., Proc. Natl. Acad. Sci. USA 84: 214-218 (1987); Nishimura et al., Canc. Res. 47: 999-1005 (1987); Wood et al., Nature 314: 446-449 (1985); and Shaw et al., J. Natl. Cancer Inst. 80: 1553-1559 (1988); Morrison, S. L., Science 229: 1202-1207 (1985); Oi et al., BioTechniques 4: 214 (1986); Winter U.S. Patent 5,225,539; Jones et al., Nature 321: 552-525 (1986); Verhoeyan et al., Science 239: 1534; and Beidler et al., J. Immunol. 141: 4053-4060 (1988).

[0129] Completely human antibodies are particularly desirable for therapeutic treatment of human patients. Such antibodies can be produced using transgenic mice that are incapable of expressing endogenous immunoglobulin heavy and light chains genes, but which can express human heavy and light chain genes. See, for example, Lonberg and Huszar, Int. Rev. Immunol. 13: 65-93 (1995); and U.S.

Patent Nos. 5,625,126; 5,633,425; 5,569,825; 5,661,016; and 5,545,806. In addition, companies such as Abgenix, Inc. (Fremont, CA) and Medarex, Inc. (Princeton, NJ), can be engaged to provide human antibodies directed against a selected antigen using technology similar to that described above. Completely human antibodies that recognize a selected epitope also can be generated using a technique referred to as "guided selection." In this approach a selected non-human monoclonal antibody (e.g., a murine antibody) is used to guide the selection of a completely human antibody recognizing the same epitope. This technology is described for example by Jespers et al., Bio/Technology 12: 899-903 (1994).

[0130] An antibody can be a single chain antibody. A single chain antibody (scFV) can be engineered (see, e.g., Colcher et al., Ann. N Y Acad. Sci. 880: 263-80 (1999); and Reiter, Clin. Cancer Res. 2: 245-52 (1996)). Single chain antibodies can be dimerized or multimerized to generate multivalent antibodies having specificities for different epitopes of the same target polypeptide.

[0131] Antibodies also may be selected or modified so that they exhibit reduced or no ability to bind an Fc receptor. For example, an antibody may be an isotype or subtype, fragment or other mutant, which does not support binding to an Fc receptor (e.g., it has a mutagenized or deleted Fc receptor binding region).

[0132] Also, an antibody (or fragment thereof) may be conjugated to a therapeutic moiety such as a cytotoxin, a therapeutic agent or a radioactive metal ion. A cytotoxin or cytotoxic agent includes any agent that is detrimental to cells. Examples include taxol, cytochalasin B, gramicidin D, ethidium bromide, emetine, mitomycin, etoposide, tenoposide, vincristine, vinblastine, colchicin, doxorubicin, daunorubicin, dihydroxy anthracin dione, mitoxantrone, mithramycin, actinomycin D, 1 dehydrotestosterone, glucocorticoids, procaine, tetracaine, lidocaine, propranolol, and puromycin and analogs or homologs thereof. Therapeutic agents include, but are not limited to, antimetabolites (e.g., methotrexate, 6-mercaptopurine, 6-thioguanine, cytarabine, 5-fluorouracil decarbazine), alkylating agents (e.g., mechlorethamine, thiotepa chlorambucil, melphalan, carmustine (BCNU) and lomustine (CCNU), cyclophosphamide, busulfan, dibromomannitol, streptozotocin, mitomycin C, and cis-dichlorodiamine platinum (II) (DDP) cisplatin), anthracyclines (e.g., daunorubicin (formerly daunomycin) and doxorubicin), antibiotics (e.g., dactinomycin (formerly actinomycin), bleomycin, mithramycin, and anthramycin (AMC)), and anti-mitotic agents (e.g., vincristine and vinblastine).

[0133] Antibody conjugates can be used for modifying a given biological response. For example, the drug moiety may be a protein or polypeptide possessing a desired biological activity. Such proteins may include, for example, a toxin such as abrin, ricin A, pseudomonas exotoxin, or diphtheria toxin; a polypeptide such as tumor necrosis factor, gamma-interferon, alpha-interferon, nerve growth factor, platelet derived growth factor, tissue plasminogen activator; or, biological response modifiers such as, for example, lymphokines, interleukin-1 ("IL-1"), interleukin-2 ("IL-2"), interleukin-6 ("IL-6"), granulocyte macrophage colony stimulating factor ("GM-CSF"), granulocyte colony stimulating factor ("G-CSF"), or

other growth factors. Also, an antibody can be conjugated to a second antibody to form an antibody heteroconjugate as described by Segal in U.S. Patent No. 4,676,980, for example.

[0134] An antibody (e.g., monoclonal antibody) can be used to isolate target polypeptides by standard techniques, such as affinity chromatography or immunoprecipitation. Moreover, an antibody can be used to detect a target polypeptide (e.g., in a cellular lysate or cell supernatant) in order to evaluate the abundance and pattern of expression of the polypeptide. Antibodies can be used diagnostically to monitor polypeptide levels in tissue as part of a clinical testing procedure, e.g., to determine the efficacy of a given treatment regimen. Detection can be facilitated by coupling (i.e., physically linking) the antibody to a detectable substance (i.e., antibody labeling). Examples of detectable substances include various enzymes, prosthetic groups, fluorescent materials, luminescent materials, bioluminescent materials, and radioactive materials. Examples of suitable enzymes include horseradish peroxidase, alkaline phosphatase, ß-galactosidase, or acetylcholinesterase; examples of suitable prosthetic group complexes include streptavidin/biotin and avidin/biotin; examples of suitable fluorescent materials include umbelliferone, fluorescein, fluorescein isothiocyanate, rhodamine, dichlorotriazinylamine fluorescein, dansyl chloride or phycoerythrin; an example of a luminescent material includes luminol; examples of bioluminescent materials include luciferase, luciferin, and aequorin, and examples of suitable radioactive material include ^{125}I, ^{131}I, ^{35}S or ^{3}H. Also, an antibody can be utilized as a test molecule for determining whether it can treat osteoarthritis, and as a therapeutic for administration to a subject for treating osteoarthritis.

[0135] An antibody can be made by immunizing with a purified antigen, or a fragment thereof, e.g., a fragment described herein, a membrane associated antigen, tissues, e.g., crude tissue preparations, whole cells, preferably living cells, lysed cells, or cell fractions.

[0136] Included herein are antibodies which bind only a native polypeptide, only denatured or otherwise non-native polypeptide, or which bind both, as well as those having linear or conformational epitopes. Conformational epitopes sometimes can be identified by selecting antibodies that bind to native but not denatured polypeptide. Also featured are antibodies that specifically bind to a polypeptide variant associated with osteoarthritis.

Methods for Identifying Candidate Therapeutics for Treating Osteoarthritis

[0137] Current therapies for the treatment of osteoarthritis have limited efficacy, limited tolerability and significant mechanism-based side effects, and few of the available therapies adequately address underlying defects. Current therapeutic approaches were largely developed in the absence of defined molecular targets or even a solid understanding of disease pathogenesis. Therefore, provided are methods of identifying candidate therapeutics that target biochemical pathways related to the development of osteoarthritis.

[0138] Thus, featured herein are methods for identifying a candidate therapeutic for treating osteoarthritis. The methods comprise contacting a test molecule with a target molecule in a system. A "target molecule" as used herein refers to a *APOL3* nucleic acid, a substantially identical nucleic acid thereof, or a fragment thereof, and an encoded polypeptide of the foregoing. The methods also comprise determining the presence or absence of an interaction between the test molecule and the target molecule, where the presence of an interaction between the test molecule and the nucleic acid or polypeptide identifies the test molecule as a candidate osteoarthritis therapeutic. The interaction between the test molecule and the target molecule may be quantified.

[0139] Test molecules and candidate therapeutics include, but are not limited to, compounds, antisense nucleic acids, siRNA molecules, ribozymes, polypeptides or proteins encoded by a *APOL3* nucleotide sequence, or a substantially identical sequence or fragment thereof, and immunotherapeutics (e.g., antibodies and HLA-presented polypeptide fragments). A test molecule or candidate therapeutic may act as a modulator of target molecule concentration or target molecule function in a system. A "modulator" may agonize (i.e., up-regulates) or antagonize (i.e., down-regulates) a target molecule concentration partially or completely in a system by affecting such cellular functions as DNA replication and/or DNA processing (e.g., DNA methylation or DNA repair), RNA transcription and/or RNA processing (e.g., removal of intronic sequences and/or translocation of spliced mRNA from the nucleus), polypeptide production (e.g., translation of the polypeptide from mRNA), and/or polypeptide post-translational modification (e.g., glycosylation, phosphorylation, and proteolysis of pro-polypeptides). A modulator may also agonize or antagonize a biological function of a target molecule partially or completely, where the function may include adopting a certain structural conformation, interacting with one or more binding partners, ligand binding, catalysis (e.g., phosphorylation, dephosphorylation, hydrolysis, methylation, and isomerization), and an effect upon a cellular event (*e.g.*, effecting progression of osteoarthritis).

[0140] As used herein, the term "system" refers to a cell free *in vitro* environment and a cell-based environment such as a collection of cells, a tissue, an organ, or an organism. A system is "contacted" with a test molecule in a variety of manners, including adding molecules in solution and allowing them to interact with one another by diffusion, cell injection, and any administration routes in an animal. As used herein, the term "interaction" refers to an effect of a test molecule on test molecule, where the effect sometimes is binding between the test molecule and the target molecule, and sometimes is an observable change in cells, tissue, or organism.

[0141] There are many standard methods for detecting the presence or absence of interaction between a test molecule and a target molecule. For example, titrametric, acidimetric, radiometric, NMR, monolayer, polarographic, spectrophotometric, fluorescent, and ESR assays probative of a target molecule interaction may be utilized.

[0142] Test molecule/target molecule interactions can be detected and/or quantified using assays known in the art. For example, an interaction can be determined by labeling the test molecule and/or the target molecule, where the label is covalently or non-covalently attached to the test molecule or target molecule. The label is sometimes a radioactive molecule such as ^{125}I, ^{131}I, ^{35}S or ^{3}H, which can be detected by direct counting of radioemission or by scintillation counting. Also, enzymatic labels such as horseradish peroxidase, alkaline phosphatase, or luciferase may be utilized where the enzymatic label can be detected by determining conversion of an appropriate substrate to product. In addition, presence or absence of an interaction can be determined without labeling. For example, a microphysiometer (*e.g.*, Cytosensor) is an analytical instrument that measures the rate at which a cell acidifies its environment using a light-addressable potentiometric sensor (LAPS). Changes in this acidification rate can be used as an indication of an interaction between a test molecule and target molecule (McConnell, H. M. *et al.*, *Science 257:* 1906-1912 (1992)).

[0143] In cell-based systems, cells typically include a *APOL3* nucleic acid, an encoded polypeptide, or substantially identical nucleic acid or polypeptide thereof, and are often of mammalian origin, although the cell can be of any origin. Whole cells, cell homogenates, and cell fractions (*e.g.*, cell membrane fractions) can be subjected to analysis. Where interactions between a test molecule with a target polypeptide are monitored, soluble and/or membrane bound forms of the polypeptide may be utilized. Where membrane-bound forms of the polypeptide are used, it may be desirable to utilize a solubilizing agent. Examples of such solubilizing agents include non-ionic detergents such as n-octylglucoside, n-dodecylglucoside, n-dodecylmaltoside, octanoyl-N-methylglucamide, decanoyl-N-methylglucamide, Triton® X-100, Triton® X-114, Thesit®, Isotridecypoly(ethylene glycol ether)$_n$, 3-[(3-cholamidopropyl)dimethylamminio]-1-propane sulfonate (CHAPS), 3-[(3-cholamidopropyl)dimethylamminio]-2-hydroxy-1-propane sulfonate (CHAPSO), or N-dodecyl-N,N-dimethyl-3-ammonio-1-propane sulfonate.

[0144] An interaction between a test molecule and target molecule also can be detected by monitoring fluorescence energy transfer (FET) (*see, e.g.*, Lakowicz *et al.*, U.S. Patent No. 5,631,169; Stavrianopoulos *et al.* U.S. Patent No. 4,868,103). A fluorophore label on a first, "donor" molecule is selected such that its emitted fluorescent energy will be absorbed by a fluorescent label on a second, "acceptor" molecule, which in turn is able to fluoresce due to the absorbed energy. Alternately, the "donor" polypeptide molecule may simply utilize the natural fluorescent energy of tryptophan residues. Labels are chosen that emit different wavelengths of light, such that the "acceptor" molecule label may be differentiated from that of the "donor". Since the efficiency of energy transfer between the labels is related to the distance separating the molecules, the spatial relationship between the molecules can be assessed. In a situation in which binding occurs between the molecules, the fluorescent emission of the

"acceptor" molecule label in the assay should be maximal. An FET binding event can be conveniently measured through standard fluorometric detection means well known in the art (*e.g.*, using a fluorimeter).

[0145] In another embodiment, determining the presence or absence of an interaction between a test molecule and a target molecule can be effected by monitoring surface plasmon resonance (*see, e.g.*, Sjolander & Urbaniczk, *Anal. Chem. 63:* 2338-2345 (1991) and Szabo *et al., Curr. Opin. Struct. Biol. 5:* 699-705 (1995)). "Surface plasmon resonance" or "biomolecular interaction analysis (BIA)" can be utilized to detect biospecific interactions in real time, without labeling any of the interactants (*e.g.*, BIAcore). Changes in the mass at the binding surface (indicative of a binding event) result in alterations of the refractive index of light near the surface (the optical phenomenon of surface plasmon resonance (SPR)), resulting in a detectable signal which can be used as an indication of real-time reactions between biological molecules.

[0146] In another embodiment, the target molecule or test molecules are anchored to a solid phase, facilitating the detection of target molecule/test molecule complexes and separation of the complexes from free, uncomplexed molecules. The target molecule or test molecule is immobilized to the solid support. In an embodiment, the target molecule is anchored to a solid surface, and the test molecule, which is not anchored, can be labeled, either directly or indirectly, with detectable labels discussed herein.

[0147] It may be desirable to immobilize a target molecule, an anti-target molecule antibody, and/or test molecules to facilitate separation of target molecule/test molecule complexes from uncomplexed forms, as well as to accommodate automation of the assay. The attachment between a test molecule and/or target molecule and the solid support may be covalent or non-covalent (*see, e.g.*, U.S. Patent No. 6,022,688 for non-covalent attachments). The solid support may be one or more surfaces of the system, such as one or more surfaces in each well of a microtiter plate, a surface of a silicon wafer, a surface of a bead (*see, e.g.*, Lam, *Nature 354:* 82-84 (1991)) that is optionally linked to another solid support, or a channel in a microfluidic device, for example. Types of solid supports, linker molecules for covalent and non-covalent attachments to solid supports, and methods for immobilizing nucleic acids and other molecules to solid supports are well known (*see, e.g.*, U.S. Patent Nos. 6,261,776; 5,900,481; 6,133,436; and 6,022,688; and WIPO publication WO 01/18234).

[0148] In an embodiment, target molecule may be immobilized to surfaces via biotin and streptavidin. For example, biotinylated target polypeptide can be prepared from biotin-NHS (N-hydroxy-succinimide) using techniques known in the art (*e.g.*, biotinylation kit, Pierce Chemicals, Rockford, IL), and immobilized in the wells of streptavidin-coated 96 well plates (Pierce Chemical). In another embodiment, a target polypeptide can be prepared as a fusion polypeptide. For example, glutathione-S-transferase/target polypeptide fusion can be adsorbed onto glutathione sepharose beads (Sigma Chemical, St. Louis, MO) or glutathione derivitized microtiter plates, which are then combined with a test molecule

under conditions conducive to complex formation (*e.g.*, at physiological conditions for salt and pH). Following incubation, the beads or microtiter plate wells are washed to remove any unbound components, or the matrix is immobilized in the case of beads, and complex formation is determined directly or indirectly as described above. Alternatively, the complexes can be dissociated from the matrix, and the level of target molecule binding or activity is determined using standard techniques.

[0149] In an embodiment, the non-immobilized component is added to the coated surface containing the anchored component. After the reaction is complete, unreacted components are removed (*e.g.*, by washing) under conditions such that a significant percentage of complexes formed will remain immobilized to the solid surface. The detection of complexes anchored on the solid surface can be accomplished in a number of manners. Where the previously non-immobilized component is pre-labeled, the detection of label immobilized on the surface indicates that complexes were formed. Where the previously non-immobilized component is not pre-labeled, an indirect label can be used to detect complexes anchored on the surface, *e.g.*, by adding a labeled antibody specific for the immobilized component, where the antibody, in turn, can be directly labeled or indirectly labeled with, *e.g.*, a labeled anti-Ig antibody.

[0150] In another embodiment, an assay is performed utilizing antibodies that specifically bind target molecule or test molecule but do not interfere with binding of the target molecule to the test molecule. Such antibodies can be derivitized to a solid support, and unbound target molecule may be immobilized by antibody conjugation. Methods for detecting such complexes, in addition to those described above for the GST-immobilized complexes, include immunodetection of complexes using antibodies reactive with the target molecule, as well as enzyme-linked assays which rely on detecting an enzymatic activity associated with the target molecule.

[0151] Cell free assays also can be conducted in a liquid phase. In such an assay, reaction products are separated from unreacted components, by any of a number of standard techniques, including but not limited to: differential centrifugation (*see, e.g.*, Rivas, G., and Minton, *Trends Biochem Sci Aug;18(8):* 284-7 (1993)); chromatography (gel filtration chromatography, ion-exchange chromatography); electrophoresis (*see, e.g.*, Ausubel *et al., eds. Current Protocols in Molecular Biology , J. Wiley:* New York (1999)); and immunoprecipitation (*see, e.g.*, Ausubel *et al., eds., supra*). Media and chromatographic techniques are known to one skilled in the art (*see, e.g.*, Heegaard, *J Mol. Recognit. Winter; 11(1-6):* 141-8 (1998); Hage & Tweed, *J. Chromatogr. B Biomed. Sci. Appl. Oct 10; 699 (1-2): 499-525* (1997)). Further, fluorescence energy transfer may also be conveniently utilized, as described herein, to detect binding without further purification of the complex from solution.

[0152] In another embodiment, modulators of target molecule expression are identified. For example, a cell or cell free mixture is contacted with a candidate compound and the expression of target mRNA or target polypeptide is evaluated relative to the level of expression of target mRNA or target

polypeptide in the absence of the candidate compound. When expression of target mRNA or target polypeptide is greater in the presence of the candidate compound than in its absence, the candidate compound is identified as an agonist of target mRNA or target polypeptide expression. Alternatively, when expression of target mRNA or target polypeptide is less (*e.g.*, less with statistical significance) in the presence of the candidate compound than in its absence, the candidate compound is identified as an antagonist or inhibitor of target mRNA or target polypeptide expression. The level of target mRNA or target polypeptide expression can be determined by methods described herein.

[0153] In another embodiment, binding partners that interact with a target molecule are detected. The target molecules can interact with one or more cellular or extracellular macromolecules, such as polypeptides *in vivo*, and these interacting molecules are referred to herein as "binding partners." Binding partners can agonize or antagonize target molecule biological activity. Also, test molecules that agonize or antagonize interactions between target molecules and binding partners can be useful as therapeutic molecules as they can up-regulate or down-regulated target molecule activity *in vivo* and thereby treat osteoarthritis.

[0154] Binding partners of target molecules can be identified by methods known in the art. For example, binding partners may be identified by lysing cells and analyzing cell lysates by electrophoretic techniques. Alternatively, a two-hybrid assay or three-hybrid assay can be utilized (*see, e.g.*, U.S. Patent No. 5,283,317; Zervos *et al., Cell 72:223-232* (1993); Madura *et al., J. Biol. Chem. 268:* 12046-12054 (1993); Bartel *et al., Biotechniques 14:* 920-924 (1993); Iwabuchi *et al., Oncogene 8:* 1693-1696 (1993); and Brent WO94/10300). A two-hybrid system is based on the modular nature of most transcription factors, which consist of separable DNA-binding and activation domains. The assay often utilizes two different DNA constructs. In one construct, a *APOL3* nucleic acid (sometimes referred to as the "bait") is fused to a gene encoding the DNA binding domain of a known transcription factor (*e.g.*, GAL-4). In another construct, a DNA sequence from a library of DNA sequences that encodes a potential binding partner (sometimes referred to as the "prey") is fused to a gene that encodes an activation domain of the known transcription factor. Sometimes, a *APOL3* nucleic acid can be fused to the activation domain. If the "bait" and the "prey" molecules interact *in vivo,* the DNA-binding and activation domains of the transcription factor are brought into close proximity. This proximity allows transcription of a reporter gene (*e.g.*, LacZ) which is operably linked to a transcriptional regulatory site responsive to the transcription factor. Expression of the reporter gene can be detected and cell colonies containing the functional transcription factor can be isolated and used to identify the potential binding partner.

[0155] In an embodiment for identifying test molecules that antagonize or agonize complex formation between target molecules and binding partners, a reaction mixture containing the target molecule and the binding partner is prepared, under conditions and for a time sufficient to allow complex formation. The reaction mixture often is provided in the presence or absence of the test molecule. The

test molecule can be included initially in the reaction mixture, or can be added at a time subsequent to the addition of the target molecule and its binding partner. Control reaction mixtures are incubated without the test molecule or with a placebo. Formation of any complexes between the target molecule and the binding partner then is detected. Decreased formation of a complex in the reaction mixture containing test molecule as compared to in a control reaction mixture indicates that the molecule antagonizes target molecule/binding partner complex formation. Alternatively, increased formation of a complex in the reaction mixture containing test molecule as compared to in a control reaction mixture indicates that the molecule agonizes target molecule/binding partner complex formation. In another embodiment, complex formation of target molecule/binding partner can be compared to complex formation of mutant target molecule/binding partner (*e.g.*, amino acid modifications in a target polypeptide). Such a comparison can be important in those cases where it is desirable to identify test molecules that modulate interactions of mutant but not non-mutated target gene products.

[0156] The assays can be conducted in a heterogeneous or homogeneous format. In heterogeneous assays, target molecule and/or the binding partner are immobilized to a solid phase, and complexes are detected on the solid phase at the end of the reaction. In homogeneous assays, the entire reaction is carried out in a liquid phase. In either approach, the order of addition of reactants can be varied to obtain different information about the molecules being tested. For example, test compounds that agonize target molecule/binding partner interactions can be identified by conducting the reaction in the presence of the test molecule in a competition format. Alternatively, test molecules that agonize preformed complexes, *e.g.*, molecules with higher binding constants that displace one of the components from the complex, can be tested by adding the test compound to the reaction mixture after complexes have been formed.

[0157] In a heterogeneous assay embodiment, the target molecule or the binding partner is anchored onto a solid surface (*e.g.*, a microtiter plate), while the non-anchored species is labeled, either directly or indirectly. The anchored molecule can be immobilized by non-covalent or covalent attachments. Alternatively, an immobilized antibody specific for the molecule to be anchored can be used to anchor the molecule to the solid surface. The partner of the immobilized species is exposed to the coated surface with or without the test molecule. After the reaction is complete, unreacted components are removed (*e.g.*, by washing) such that a significant portion of any complexes formed will remain immobilized on the solid surface. Where the non-immobilized species is pre-labeled, the detection of label immobilized on the surface is indicative of complex. Where the non-immobilized species is not pre-labeled, an indirect label can be used to detect complexes anchored to the surface; *e.g.*, by using a labeled antibody specific for the initially non-immobilized species. Depending upon the order of addition of reaction components, test compounds that inhibit complex formation or that disrupt preformed complexes can be detected.

[0158] In another embodiment, the reaction can be conducted in a liquid phase in the presence or absence of test molecule, where the reaction products are separated from unreacted components, and the complexes are detected (*e.g.*, using an immobilized antibody specific for one of the binding components to anchor any complexes formed in solution, and a labeled antibody specific for the other partner to detect anchored complexes). Again, depending upon the order of addition of reactants to the liquid phase, test compounds that inhibit complex or that disrupt preformed complexes can be identified.

[0159] In an alternate embodiment, a homogeneous assay can be utilized. For example, a preformed complex of the target gene product and the interactive cellular or extracellular binding partner product is prepared. One or both of the target molecule or binding partner is labeled, and the signal generated by the label(s) is quenched upon complex formation (*e.g.*, U.S. Patent No. 4,109,496 that utilizes this approach for immunoassays). Addition of a test molecule that competes with and displaces one of the species from the preformed complex will result in the generation of a signal above background. In this way, test substances that disrupt target molecule/binding partner complexes can be identified.

[0160] Candidate therapeutics for treating osteoarthritis are identified from a group of test molecules that interact with a target molecule. Test molecules are normally ranked according to the degree with which they modulate (*e.g.*, agonize or antagonize) a function associated with the target molecule (*e.g.*, DNA replication and/or processing, RNA transcription and/or processing, polypeptide production and/or processing, and/or biological function/activity), and then top ranking modulators are selected. Also, pharmacogenomic information described herein can determine the rank of a modulator. The top 10% of ranked test molecules often are selected for further testing as candidate therapeutics, and sometimes the top 15%, 20%, or 25% of ranked test molecules are selected for further testing as candidate therapeutics. Candidate therapeutics typically are formulated for administration to a subject.

Therapeutic Formulations

[0161] Formulations and pharmaceutical compositions typically include in combination with a pharmaceutically acceptable carrier one or more target molecule modulators. The modulator often is a test molecule identified as having an interaction with a target molecule by a screening method described above. The modulator may be a compound, an antisense nucleic acid, a ribozyme, an antibody, or a binding partner. Also, formulations may comprise a target polypeptide or fragment thereof in combination with a pharmaceutically acceptable carrier, where the polypeptide or fragment sometimes has an *APOL3* biological activity (e.g., apolipoprotein activity), and sometimes includes all or part of an apolipoprotein domain.

[0162] As used herein, the term "pharmaceutically acceptable carrier" includes solvents, dispersion media, coatings, antibacterial and antifungal agents, isotonic and absorption delaying agents, and the like, compatible with pharmaceutical administration. Supplementary active compounds can also be

incorporated into the compositions. Pharmaceutical compositions can be included in a container, pack, or dispenser together with instructions for administration.

[0163] A pharmaceutical composition typically is formulated to be compatible with its intended route of administration. Examples of routes of administration include parenteral, *e.g.*, intravenous, intradermal, subcutaneous, oral (*e.g.*, inhalation), transdermal (topical), transmucosal, and rectal administration. Solutions or suspensions used for parenteral, intradermal, or subcutaneous application can include the following components: a sterile diluent such as water for injection, saline solution, fixed oils, polyethylene glycols, glycerin, propylene glycol or other synthetic solvents; antibacterial agents such as benzyl alcohol or methyl parabens; antioxidants such as ascorbic acid or sodium bisulfite; chelating agents such as ethylenediaminetetraacetic acid; buffers such as acetates, citrates or phosphates and agents for the adjustment of tonicity such as sodium chloride or dextrose. pH can be adjusted with acids or bases, such as hydrochloric acid or sodium hydroxide. The parenteral preparation can be enclosed in ampoules, disposable syringes or multiple dose vials made of glass or plastic.

[0164] Oral compositions generally include an inert diluent or an edible carrier. For the purpose of oral therapeutic administration, the active compound can be incorporated with excipients and used in the form of tablets, troches, or capsules, *e.g.*, gelatin capsules. Oral compositions can also be prepared using a fluid carrier for use as a mouthwash. Pharmaceutically compatible binding agents, and/or adjuvant materials can be included as part of the composition. The tablets, pills, capsules, troches and the like can contain any of the following ingredients, or compounds of a similar nature: a binder such as microcrystalline cellulose, gum tragacanth or gelatin; an excipient such as starch or lactose, a disintegrating agent such as alginic acid, Primogel, or corn starch; a lubricant such as magnesium stearate or Sterotes; a glidant such as colloidal silicon dioxide; a sweetening agent such as sucrose or saccharin; or a flavoring agent such as peppermint, methyl salicylate, or orange flavoring.

[0165] Pharmaceutical compositions suitable for injectable use include sterile aqueous solutions (where water soluble) or dispersions and sterile powders for the extemporaneous preparation of sterile injectable solutions or dispersion. For intravenous administration, suitable carriers include physiological saline, bacteriostatic water, Cremophor EL™ (BASF, Parsippany, NJ) or phosphate buffered saline (PBS). In all cases, the composition must be sterile and should be fluid to the extent that easy syringability exists. It should be stable under the conditions of manufacture and storage and must be preserved against the contaminating action of microorganisms such as bacteria and fungi. The carrier can be a solvent or dispersion medium containing, for example, water, ethanol, polyol (for example, glycerol, propylene glycol, and liquid polyethylene glycol, and the like), and suitable mixtures thereof. The proper fluidity can be maintained, for example, by the use of a coating such as lecithin, by the maintenance of the required particle size in the case of dispersion and by the use of surfactants. Prevention of the action of microorganisms can be achieved by various antibacterial and antifungal agents, for example, parabens,

chlorobutanol, phenol, ascorbic acid, thimerosal, and the like. In many cases, it will be preferable to include isotonic agents, for example, sugars, polyalcohols such as mannitol, sorbitol, sodium chloride in the composition. Prolonged absorption of the injectable compositions can be brought about by including in the composition an agent which delays absorption, for example, aluminum monostearate and gelatin.

[0166] Sterile injectable solutions can be prepared by incorporating the active compound in the required amount in an appropriate solvent with one or a combination of ingredients enumerated above, as required, followed by filtered sterilization. Generally, dispersions are prepared by incorporating the active compound into a sterile vehicle which contains a basic dispersion medium and the required other ingredients from those enumerated above. In the case of sterile powders for the preparation of sterile injectable solutions, the preferred methods of preparation are vacuum drying and freeze-drying which yields a powder of the active ingredient plus any additional desired ingredient from a previously sterile-filtered solution thereof.

[0167] For administration by inhalation, the compounds are delivered in the form of an aerosol spray from pressured container or dispenser which contains a suitable propellant, *e.g.*, a gas such as carbon dioxide, or a nebulizer.

[0168] Systemic administration can also be by transmucosal or transdermal means. For transmucosal or transdermal administration, penetrants appropriate to the barrier to be permeated are used in the formulation. Such penetrants are generally known in the art, and include, for example, for transmucosal administration, detergents, bile salts, and fusidic acid derivatives. Transmucosal administration can be accomplished through the use of nasal sprays or suppositories. For transdermal administration, the active compounds are formulated into ointments, salves, gels, or creams as generally known in the art. Molecules can also be prepared in the form of suppositories (*e.g.*, with conventional suppository bases such as cocoa butter and other glycerides) or retention enemas for rectal delivery.

[0169] In one embodiment, active molecules are prepared with carriers that will protect the compound against rapid elimination from the body, such as a controlled release formulation, including implants and microencapsulated delivery systems. Biodegradable, biocompatible polymers can be used, such as ethylene vinyl acetate, polyanhydrides, polyglycolic acid, collagen, polyorthoesters, and polylactic acid. Methods for preparation of such formulations will be apparent to those skilled in the art. Materials can also be obtained commercially from Alza Corporation and Nova Pharmaceuticals, Inc. Liposomal suspensions (including liposomes targeted to infected cells with monoclonal antibodies to viral antigens) can also be used as pharmaceutically acceptable carriers. These can be prepared according to methods known to those skilled in the art, for example, as described in U.S. Patent No. 4,522,811.

[0170] It is advantageous to formulate oral or parenteral compositions in dosage unit form for ease of administration and uniformity of dosage. Dosage unit form as used herein refers to physically discrete

units suited as unitary dosages for the subject to be treated; each unit containing a predetermined quantity of active compound calculated to produce the desired therapeutic effect in association with the required pharmaceutical carrier.

[0171] Toxicity and therapeutic efficacy of such compounds can be determined by standard pharmaceutical procedures in cell cultures or experimental animals, *e.g.,* for determining the LD_{50} (the dose lethal to 50% of the population) and the ED_{50} (the dose therapeutically effective in 50% of the population). The dose ratio between toxic and therapeutic effects is the therapeutic index and it can be expressed as the ratio LD_{50}/ED_{50}. Molecules which exhibit high therapeutic indices are preferred. While molecules that exhibit toxic side effects may be used, care should be taken to design a delivery system that targets such compounds to the site of affected tissue in order to minimize potential damage to uninfected cells and, thereby, reduce side effects.

[0172] The data obtained from the cell culture assays and animal studies can be used in formulating a range of dosage for use in humans. The dosage of such molecules lies preferably within a range of circulating concentrations that include the ED_{50} with little or no toxicity. The dosage may vary within this range depending upon the dosage form employed and the route of administration utilized. For any molecules used in the methods described herein, the therapeutically effective dose can be estimated initially from cell culture assays. A dose may be formulated in animal models to achieve a circulating plasma concentration range that includes the IC_{50} (*i.e.*, the concentration of the test compound which achieves a half-maximal inhibition of symptoms) as determined in cell culture. Such information can be used to more accurately determine useful doses in humans. Levels in plasma may be measured, for example, by high performance liquid chromatography.

[0173] As defined herein, a therapeutically effective amount of protein or polypeptide (*i.e.*, an effective dosage) ranges from about 0.001 to 30 mg/kg body weight, sometimes about 0.01 to 25 mg/kg body weight, often about 0.1 to 20 mg/kg body weight, and more often about 1 to 10 mg/kg, 2 to 9 mg/kg, 3 to 8 mg/kg, 4 to 7 mg/kg, or 5 to 6 mg/kg body weight. The protein or polypeptide can be administered one time per week for between about 1 to 10 weeks, sometimes between 2 to 8 weeks, often between about 3 to 7 weeks, and more often for about 4, 5, or 6 weeks. The skilled artisan will appreciate that certain factors may influence the dosage and timing required to effectively treat a subject, including but not limited to the severity of the disease or disorder, previous treatments, the general health and/or age of the subject, and other diseases present. Moreover, treatment of a subject with a therapeutically effective amount of a protein, polypeptide, or antibody can include a single treatment or, preferably, can include a series of treatments.

[0174] With regard to polypeptide formulations, featured herein is a method for treating osteoarthritis in a subject, which comprises contacting one or more cells in the subject with a first polypeptide, where the subject comprises a second polypeptide having one or more polymorphic

variations associated with cancer, and where the first polypeptide comprises fewer polymorphic variations associated with cancer than the second polypeptide. The first and second polypeptides are encoded by a nucleic acid which comprises a nucleotide sequence in SEQ ID NO: 1-7; a nucleotide sequence which encodes a polypeptide consisting of an amino acid sequence encoded by a nucleotide sequence referenced in SEQ ID NO: 1-7; a nucleotide sequence which encodes a polypeptide that is 90% or more identical to an amino acid sequence encoded by a nucleotide sequence of SEQ ID NO: 1-7 and a nucleotide sequence 90% or more identical to a nucleotide sequence in SEQ ID NO: 1-7. The subject often is a human.

[0175] For antibodies, a dosage of 0.1 mg/kg of body weight (generally 10 mg/kg to 20 mg/kg) is often utilized. If the antibody is to act in the brain, a dosage of 50 mg/kg to 100 mg/kg is often appropriate. Generally, partially human antibodies and fully human antibodies have a longer half-life within the human body than other antibodies. Accordingly, lower dosages and less frequent administration is often possible. Modifications such as lipidation can be used to stabilize antibodies and to enhance uptake and tissue penetration (*e.g.*, into the brain). A method for lipidation of antibodies is described by Cruikshank *et al.*, *J. Acquired Immune Deficiency Syndromes and Human Retrovirology 14:*193 (1997).

[0176] Antibody conjugates can be used for modifying a given biological response, the drug moiety is not to be construed as limited to classical chemical therapeutic agents. For example, the drug moiety may be a protein or polypeptide possessing a desired biological activity. Such proteins may include, for example, a toxin such as abrin, ricin A, pseudomonas exotoxin, or diphtheria toxin; a polypeptide such as tumor necrosis factor, alpha-interferon, beta-interferon, nerve growth factor, platelet derived growth factor, tissue plasminogen activator; or, biological response modifiers such as, for example, lymphokines, interleukin-1 ("IL-1"), interleukin-2 ("IL-2"), interleukin-6 ("IL-6"), granulocyte macrophage colony stimulating factor ("GM-CSF"), granulocyte colony stimulating factor ("G-CSF"), or other growth factors. Alternatively, an antibody can be conjugated to a second antibody to form an antibody heteroconjugate as described by Segal in U.S. Patent No. 4,676,980.

[0177] For compounds, exemplary doses include milligram or microgram amounts of the compound per kilogram of subject or sample weight, for example, about 1 microgram per kilogram to about 500 milligrams per kilogram, about 100 micrograms per kilogram to about 5 milligrams per kilogram, or about 1 microgram per kilogram to about 50 micrograms per kilogram. It is understood that appropriate doses of a small molecule depend upon the potency of the small molecule with respect to the expression or activity to be modulated. When one or more of these small molecules is to be administered to an animal (*e.g.*, a human) in order to modulate expression or activity of a polypeptide or nucleic acid described herein, a physician, veterinarian, or researcher may, for example, prescribe a relatively low dose at first, subsequently increasing the dose until an appropriate response is obtained. In addition, it is

understood that the specific dose level for any particular animal subject will depend upon a variety of factors including the activity of the specific compound employed, the age, body weight, general health, gender, and diet of the subject, the time of administration, the route of administration, the rate of excretion, any drug combination, and the degree of expression or activity to be modulated.

[0178] With regard to nucleic acid formulations, gene therapy vectors can be delivered to a subject by, for example, intravenous injection, local administration (*see, e.g.,* U.S. Patent 5,328,470) or by stereotactic injection (*see e.g.*, Chen *et al.*, (1994) *Proc. Natl. Acad. Sci. USA 91*:3054-3057). Pharmaceutical preparations of gene therapy vectors can include a gene therapy vector in an acceptable diluent, or can comprise a slow release matrix in which the gene delivery vehicle is imbedded. Alternatively, where the complete gene delivery vector can be produced intact from recombinant cells (*e.g.*, retroviral vectors) the pharmaceutical preparation can include one or more cells which produce the gene delivery system. Examples of gene delivery vectors are described herein.

<u>Therapeutic Methods</u>

[0179] A therapeutic formulation described above can be administered to a subject in need of a therapeutic for inducing a desired biological response.. Therapeutic formulations can be administered by any of the paths described herein. With regard to both prophylactic and therapeutic methods of treatment, such treatments may be specifically tailored or modified, based on knowledge obtained from pharmacogenomic analyses described herein.

[0180] As used herein, the term "treatment" is defined as the application or administration of a therapeutic formulation to a subject, or application or administration of a therapeutic agent to an isolated tissue or cell line from a subject with the purpose to cure, heal, alleviate, relieve, alter, remedy, ameliorate, improve or affect osteoarthritis, symptoms of osteoarthritis or a predisposition towards osteoarthritis. A therapeutic formulation includes, but is not limited to, small molecules, peptides, antibodies, ribozymes and antisense oligonucleotides. Administration of a therapeutic formulation can occur prior to the manifestation of symptoms characteristic of osteoarthritis, such that osteoarthritis is prevented or delayed in its progression. The appropriate therapeutic composition can be determined based on screening assays described herein.

[0181] As discussed, successful treatment of osteoarthritis can be brought about by techniques that serve to agonize target molecule expression or function, or alternatively, antagonize target molecule expression or function. These techniques include administration of modulators that include, but are not limited to, small organic or inorganic molecules; antibodies (including, for example, polyclonal, monoclonal, humanized, anti-idiotypic, chimeric or single chain antibodies, and Fab, $F(ab')_2$ and Fab expression library fragments, scFV molecules, and epitope-binding fragments thereof); and peptides, phosphopeptides, or polypeptides.

[0182] Further, antisense and ribozyme molecules that inhibit expression of the target gene can also be used to reduce the level of target gene expression, thus effectively reducing the level of target gene activity. Still further, triple helix molecules can be utilized in reducing the level of target gene activity. Antisense, ribozyme and triple helix molecules are discussed above. It is possible that the use of antisense, ribozyme, and/or triple helix molecules to reduce or inhibit mutant gene expression can also reduce or inhibit the transcription (triple helix) and/or translation (antisense, ribozyme) of mRNA produced by normal target gene alleles, such that the concentration of normal target gene product present can be lower than is necessary for a normal phenotype. In such cases, nucleic acid molecules that encode and express target gene polypeptides exhibiting normal target gene activity can be introduced into cells via gene therapy method. Alternatively, in instances in that the target gene encodes an extracellular polypeptide, it can be preferable to co-administer normal target gene polypeptide into the cell or tissue in order to maintain the requisite level of cellular or tissue target gene activity.

[0183] Another method by which nucleic acid molecules may be utilized in treating or preventing osteoarthritis is use of aptamer molecules specific for target molecules. Aptamers are nucleic acid molecules having a tertiary structure which permits them to specifically bind to ligands (*see, e.g.*, Osborne, *et al., Curr. Opin. Chem. Biol.1(1):* 5-9 (1997); and Patel, D. J., *Curr. Opin. Chem. Biol. Jun;1(1):* 32-46 (1997)).

[0184] Yet another method of utilizing nucleic acid molecules for osteoarthritis treatment is gene therapy, which can also be referred to as allele therapy. Provided herein is a gene therapy method for treating osteoarthritis in a subject, which comprises contacting one or more cells in the subject or from the subject with a nucleic acid having a first nucleotide sequence (e.g., the first nucleotide sequence is identical to or substantially identical to the nucleotide sequence of SEQ ID NO: 1-7). Genomic DNA in the subject comprises a second nucleotide sequence having one or more polymorphic variations associated with osteoarthritis (e.g., the second nucleotide sequence is identical to or substantially identical to a nucleotide sequence of SEQ ID NO: 1). The first and second nucleotide sequences typically are substantially identical to one another, and the first nucleotide sequence comprises fewer polymorphic variations associated with osteoarthritis than the second nucleotide sequence. The first nucleotide sequence may comprise a gene sequence that encodes a full-length polypeptide or a fragment thereof. The subject is often a human. Allele therapy methods often are utilized in conjunction with a method of first determining whether a subject has genomic DNA that includes polymorphic variants associated with osteoarthritis.

[0185] In another allele therapy embodiment, provided herein is a method which comprises contacting one or more cells in the subject or from the subject with a polypeptide encoded by a nucleic acid having a first nucleotide sequence (e.g., the first nucleotide sequence is identical to or substantially identical to the nucleotide sequence of SEQ ID NO: 1-7). Genomic DNA in the subject comprises a

second nucleotide sequence having one or more polymorphic variations associated with osteoarthritis (e.g., the second nucleic acid has a nucleotide sequence identical to or substantially identical to the nucleotide sequence of SEQ ID NO: 1). The first and second nucleotide sequences typically are substantially identical to one another, and the first nucleotide sequence comprises fewer polymorphic variations associated with osteoarthritis than the second nucleotide sequence. The first nucleotide sequence may comprise a gene sequence that encodes a full-length polypeptide or a fragment thereof. The subject is often a human.

[0186] For antibody-based therapies, antibodies can be generated that are both specific for target molecules and that reduce target molecule activity. Such antibodies may be administered in instances where antagonizing a target molecule function is appropriate for the treatment of osteoarthritis.

[0187] In circumstances where stimulating antibody production in an animal or a human subject by injection with a target molecule is harmful to the subject, it is possible to generate an immune response against the target molecule by use of anti-idiotypic antibodies (*see, e.g.*, Herlyn, *Ann. Med.;31(1):* 66-78 (1999); and Bhattacharya-Chatterjee & Foon, *Cancer Treat. Res.; 94:* 51-68 (1998)). Introducing an anti-idiotypic antibody to a mammal or human subject often stimulates production of anti-anti-idiotypic antibodies, which typically are specific to the target molecule. Vaccines directed to osteoarthritis also may be generated in this fashion.

[0188] In instances where the target molecule is intracellular and whole antibodies are used, internalizing antibodies may be preferred. Lipofectin or liposomes can be used to deliver the antibody or a fragment of the Fab region that binds to the target antigen into cells. Where fragments of the antibody are used, the smallest inhibitory fragment that binds to the target antigen is preferred. For example, peptides having an amino acid sequence corresponding to the Fv region of the antibody can be used. Alternatively, single chain neutralizing antibodies that bind to intracellular target antigens can also be administered. Such single chain antibodies can be administered, for example, by expressing nucleotide sequences encoding single-chain antibodies within the target cell population (*see, e.g.*, Marasco *et al.*, *Proc. Natl. Acad. Sci. USA 90:* 7889-7893 (1993)).

[0189] Modulators can be administered to a patient at therapeutically effective doses to treat osteoarthritis. A therapeutically effective dose refers to an amount of the modulator sufficient to result in amelioration of symptoms of osteoarthritis. Toxicity and therapeutic efficacy of modulators can be determined by standard pharmaceutical procedures in cell cultures or experimental animals, *e.g.*, for determining the LD_{50} (the dose lethal to 50% of the population) and the ED_{50} (the dose therapeutically effective in 50% of the population). The dose ratio between toxic and therapeutic effects is the therapeutic index and it can be expressed as the ratio LD_{50}/ED_{50}. Modulators that exhibit large therapeutic indices are preferred. While modulators that exhibit toxic side effects can be used, care

should be taken to design a delivery system that targets such molecules to the site of affected tissue in order to minimize potential damage to uninfected cells, thereby reducing side effects.

[0190] Data obtained from cell culture assays and animal studies can be used in formulating a range of dosages for use in humans. The dosage of such compounds lies preferably within a range of circulating concentrations that include the ED_{50} with little or no toxicity. The dosage can vary within this range depending upon the dosage form employed and the route of administration utilized. For any compound used in the methods described herein, the therapeutically effective dose can be estimated initially from cell culture assays. A dose can be formulated in animal models to achieve a circulating plasma concentration range that includes the IC_{50} (*i.e.*, the concentration of the test compound that achieves a half-maximal inhibition of symptoms) as determined in cell culture. Such information can be used to more accurately determine useful doses in humans. Levels in plasma can be measured, for example, by high performance liquid chromatography.

[0191] Another example of effective dose determination for an individual is the ability to directly assay levels of "free" and "bound" compound in the serum of the test subject. Such assays may utilize antibody mimics and/or "biosensors" that have been created through molecular imprinting techniques. Molecules that modulate target molecule activity are used as a template, or "imprinting molecule", to spatially organize polymerizable monomers prior to their polymerization with catalytic reagents. The subsequent removal of the imprinted molecule leaves a polymer matrix which contains a repeated "negative image" of the compound and is able to selectively rebind the molecule under biological assay conditions. A detailed review of this technique can be seen in Ansell *et al., Current Opinion in Biotechnology 7:* 89-94 (1996) and in Shea, *Trends in Polymer Science 2:* 166-173 (1994). Such "imprinted" affinity matrixes are amenable to ligand-binding assays, whereby the immobilized monoclonal antibody component is replaced by an appropriately imprinted matrix. An example of the use of such matrixes in this way can be seen in Vlatakis, *et al., Nature 361:* 645-647 (1993). Through the use of isotope-labeling, the "free" concentration of compound which modulates target molecule expression or activity readily can be monitored and used in calculations of IC_{50}. Such "imprinted" affinity matrixes can also be designed to include fluorescent groups whose photon-emitting properties measurably change upon local and selective binding of target compound. These changes readily can be assayed in real time using appropriate fiberoptic devices, in turn allowing the dose in a test subject to be quickly optimized based on its individual IC_{50}. An example of such a "biosensor" is discussed in Kriz *et al., Analytical Chemistry 67:* 2142-2144 (1995).

[0192] The examples set forth below are intended to illustrate but not limit the invention.

Examples

[0193] In the following studies a group of subjects was selected according to specific parameters relating to osteoarthritis (OA). Nucleic acid samples obtained from individuals in the study group were subjected to genetic analysis, which identified associations between osteoarthritis and a polymorphism in the *APOL3* gene on chromosome twenty-two. The polymorphism was genotyped again in two replication cohorts consisting of individuals selected for OA. In addition, SNPs proximal to the incident polymorphism were identified and allelotyped in OA case and control pools. Methods are described for producing *APOL3* polypeptide and *APOL3* polypeptide variants *in vitro* or *in vivo*, *APOL3* nucleic acids or polypeptides and variants thereof are utilized for screening test molecules for those that interact with *APOL3* molecules. Test molecules identified as interactors with *APOL3* molecules and *APOL3* variants are further screened *in vivo* to determine whether they treat osteoarthritis.

Example 1
Samples and Pooling Strategies

Sample Selection

[0194] Blood samples were collected from individuals diagnosed with knee osteoarthritis, which were referred to as case samples. Also, blood samples were collected from individuals not diagnosed with knee osteoarthritis as gender and age-matched controls. A database was created that listed all phenotypic trait information gathered from individuals for each case and control sample. Genomic DNA was extracted from each of the blood samples for genetic analyses.

DNA Extraction from Blood Samples

[0195] Six to ten milliliters of whole blood was transferred to a 50 ml tube containing 27 ml of red cell lysis solution (RCL). The tube was inverted until the contents were mixed. Each tube was incubated for 10 minutes at room temperature and inverted once during the incubation. The tubes were then centrifuged for 20 minutes at 3000 x g and the supernatant was carefully poured off. 100-200 μl of residual liquid was left in the tube and was pipetted repeatedly to resuspend the pellet in the residual supernatant. White cell lysis solution (WCL) was added to the tube and pipetted repeatedly until completely mixed. While no incubation was normally required, the solution was incubated at 37°C or room temperature if cell clumps were visible after mixing until the solution was homogeneous. 2 ml of protein precipitation was added to the cell lysate. The mixtures were vortexed vigorously at high speed for 20 sec to mix the protein precipitation solution uniformly with the cell lysate, and then centrifuged for 10 minutes at 3000 x g. The supernatant containing the DNA was then poured into a clean 15 ml tube, which contained 7 ml of 100% isopropanol. The samples were mixed by inverting the tubes gently until

white threads of DNA were visible. Samples were centrifuged for 3 minutes at 2000 x g and the DNA was visible as a small white pellet. The supernatant was decanted and 5 ml of 70% ethanol was added to each tube. Each tube was inverted several times to wash the DNA pellet, and then centrifuged for 1 minute at 2000 x g. The ethanol was decanted and each tube was drained on clean absorbent paper. The DNA was dried in the tube by inversion for 10 minutes, and then 1000 μl of 1X TE was added. The size of each sample was estimated, and less TE buffer was added during the following DNA hydration step if the sample was smaller. The DNA was allowed to rehydrate overnight at room temperature, and DNA samples were stored at 2-8°C.

[0196] DNA was quantified by placing samples on a hematology mixer for at least 1 hour. DNA was serially diluted (typically 1:80, 1:160, 1:320, and 1:640 dilutions) so that it would be within the measurable range of standards. 125 μl of diluted DNA was transferred to a clear U-bottom microtitre plate, and 125 μl of 1X TE buffer was transferred into each well using a multichannel pipette. The DNA and 1X TE were mixed by repeated pipetting at least 15 times, and then the plates were sealed. 50 μl of diluted DNA was added to wells A5-H12 of a black flat bottom microtitre plate. Standards were inverted six times to mix them, and then 50 μl of 1X TE buffer was pipetted into well A1, 1000 ng/ml of standard was pipetted into well A2, 500 ng/ml of standard was pipetted into well A3, and 250 ng/ml of standard was pipetted into well A4. PicoGreen (Molecular Probes, Eugene, Oregon) was thawed and freshly diluted 1:200 according to the number of plates that were being measured. PicoGreen was vortexed and then 50μl was pipetted into all wells of the black plate with the diluted DNA. DNA and PicoGreen were mixed by pipetting repeatedly at least 10 times with the multichannel pipette. The plate was placed into a Fluoroskan Ascent Machine (microplate fluorometer produced by Labsystems) and the samples were allowed to incubate for 3 minutes before the machine was run using filter pairs 485 nm excitation and 538 nm emission wavelengths. Samples having measured DNA concentrations of greater than 450 ng/μl were re-measured for conformation. Samples having measured DNA concentrations of 20 ng/μl or less were re-measured for confirmation.

<u>Pooling Strategies – Discovery Cohort</u>

[0197] Samples were derived from the Nottingham knee OA family study (UK) where index cases were identified through a knee replacement registry. Siblings were approached and assessed with knee x-rays and assigned status as affected or unaffected. In all 1,157 individuals were available. In order to create same-sex pools of appropriate sizes, 335 unrelated female individuals with OA from the Nottingham OA sample were selected for the case pool. The control pool was made up of unrelated female individuals from the St. Thomas twin study (England) with normal knee x-rays and without other indications of OA, regardless of anatomical location, as well as lacking family history of OA. The St.

Thomas twin study consists of Caucasian, female participants from the St. Thomas' Hospital, London, adult-twin registry, which is a voluntary registry of >4,000 twin pairs ranging from 18 to 76 years of age. The female case samples and female control samples are described further in Table 1 below.

[0198] A select set of samples from each group were utilized to generate pools, and one pool was created for each group. Each individual sample in a pool was represented by an equal amount of genomic DNA. For example, where 25 ng of genomic DNA was utilized in each PCR reaction and there were 200 individuals in each pool, each individual would provide 125 pg of genomic DNA. Inclusion or exclusion of samples for a pool was based upon the following criteria: the sample was derived from an individual characterized as Caucasian; the sample was derived from an individual of British paternal and maternal descent; case samples were derived from individuals diagnosed with specific knee osteoarthritis (OA) and were recruited from an OA knee replacement clinic. Control samples were derived from individuals free of OA, family history of OA, and rheumatoid arthritis. Also, sufficient genomic DNA was extracted from each blood sample for all allelotyping and genotyping reactions performed during the study. Phenotype information from each individual was collected and included age of the individual, gender, family history of OA, general medical information (e.g., height, weight, thyroid disease, diabetes, psoriasis, hysterectomy), joint history (previous and current symptoms, joint-related operations, age at onset of symptoms, date of primary diagnosis, age of individual as of primary diagnosis and order of involvement), and knee-related findings (crepitus, restricted passive movement, bony swelling/deformity). Additional knee information included knee history, current symptoms, any major knee injury, menisectomy, knee replacement surgery, age of surgery, and treatment history (including hormone replace therapy (HRT)). Samples that met these criteria were added to appropriate pools based on disease status.

[0199] The selection process yielded the pools set forth in Table 1, which were used in the studies that follow:

TABLE 1

	Female case	**Female control**
Pool size (Number)	335	335
Pool Criteria (ex: case/control)	control	case
Mean Age (ex: years)	57.21	69.95

Example 2

Association of Polymorphic Variants with Osteoarthritis

[0200] A whole-genome screen was performed to identify particular SNPs associated with occurrence of osteoarthritis. As described in Example 1, two sets of samples were utilized, which included samples from female individuals having knee osteoarthritis (osteoarthritis cases), and samples from female individuals not having knee osteoarthritis (female controls). The initial screen of each pool was performed in an allelotyping study, in which certain samples in each group were pooled. By pooling DNA from each group, an allele frequency for each SNP in each group was calculated. These allele frequencies were then compared to one another. Particular SNPs were considered as being associated with osteoarthritis when allele frequency differences calculated between case and control pools were statistically significant. SNP disease association results obtained from the allelotyping study were then validated by genotyping each associated SNP across all samples from each pool. The results of the genotyping then were analyzed, allele frequencies for each group were calculated from the individual genotyping results, and a p-value was calculated to determine whether the case and control groups had statistically significant differences in allele frequencies for a particular SNP. When the genotyping results agreed with the original allelotyping results, the SNP disease association was considered validated at the genetic level.

SNP Panel Used for Genetic Analyses

[0201] A whole-genome SNP screen began with an initial screen of approximately 25,000 SNPs over each set of disease and control samples using a pooling approach. The pools studied in the screen are described in Example 1. The SNPs analyzed in this study were part of a set of 25,488 SNPs confirmed as being statistically polymorphic as each is characterized as having a minor allele frequency of greater than 10%. The SNPs in the set reside in genes or in close proximity to genes, and many reside in gene exons. Specifically, SNPs in the set are located in exons, introns, and within 5,000 base-pairs upstream of a transcription start site of a gene. In addition, SNPs were selected according to the following criteria: they are located in ESTs; they are located in Locuslink or Ensembl genes; and they are located in Genomatix promoter predictions. SNPs in the set were also selected on the basis of even spacing across the genome, as depicted in Table 2.

[0202] A case-control study design using a whole genome association strategy involving approximately 28,000 single nucleotide polymorphisms (SNPs) was employed. Approximately 25,000 SNPs were evenly spaced in gene-based regions of the human genome with a median inter-marker distance of about 40,000 base pairs. Additionally, approximately 3,000 SNPs causing amino acid substitutions in genes described in the literature as candidates for various diseases were used. The case-

control study samples were of female Caucasian origin (British paternal and maternal descent) 670 individuals were equally distributed in two groups: female controls and female cases. The whole genome association approach was first conducted on 2 DNA pools representing the 2 groups. Significant markers were confirmed by individual genotyping.

TABLE 2

General Statistics		**Spacing Statistics**	
Total # of SNPs	25,488	Median	37,058 bp
# of Exonic SNPs	>4,335 (17%)	Minimum*	1,000 bp
# SNPs with refSNP ID	20,776 (81%)	Maximum*	3,000,000 bp
Gene Coverage	>10,000	Mean	122,412 bp
Chromosome Coverage	All	Std Deviation	373,325 bp
		**Excludes outliers*	

Allelotyping and Genotyping Results

[0203] The genetic studies summarized above and described in more detail below identified an allelic variant in the APOL3 gene that is associated with osteoarthritis.

Assay for Verifying, Allelotyping, and Genotyping SNPs

[0204] A MassARRAY™ system (Sequenom, Inc.) was utilized to perform SNP genotyping in a high-throughput fashion. This genotyping platform was complemented by a homogeneous, single-tube assay method (hME™ or homogeneous MassEXTEND™ (Sequenom, Inc.)) in which two genotyping primers anneal to and amplify a genomic target surrounding a polymorphic site of interest. A third primer (the MassEXTEND™ primer), which is complementary to the amplified target up to but not including the polymorphism, was then enzymatically extended one or a few bases through the polymorphic site and then terminated.

[0205] For each polymorphism, SpectroDESIGNER™ software (Sequenom, Inc.) was used to generate a set of PCR primers and a MassEXTEND™ primer which where used to genotype the polymorphism. Other primer design software could be used or one of ordinary skill in the art could manually design primers based on his or her knowledge of the relevant factors and considerations in designing such primers. Table 3 shows PCR primers and Table 4 shows extension primers used for analyzing polymorphisms. The initial PCR amplification reaction was performed in a 5 μl total volume containing 1X PCR buffer with 1.5 mM $MgCl_2$ (Qiagen), 200 μM each of dATP, dGTP, dCTP, dTTP (Gibco-BRL), 2.5 ng of genomic DNA, 0.1 units of HotStar DNA polymerase (Qiagen), and 200 nM each of forward and reverse PCR primers specific for the polymorphic region of interest.

TABLE 3: PCR Primers

SNP Reference	Forward PCR primer	Reverse PCR primer
rs132659	ACGTTGGATGGGCCCATAGTGGGTCATAAC	ACGTTGGATGGTGGGGTGAGTGCCCAAAAG

[0206] Samples were incubated at 95°C for 15 minutes, followed by 45 cycles of 95°C for 20 seconds, 56°C for 30 seconds, and 72°C for 1 minute, finishing with a 3 minute final extension at 72°C. Following amplification, shrimp alkaline phosphatase (SAP) (0.3 units in a 2 μl volume) (Amersham Pharmacia) was added to each reaction (total reaction volume was 7 μl) to remove any residual dNTPs that were not consumed in the PCR step. Samples were incubated for 20 minutes at 37°C, followed by 5 minutes at 85°C to denature the SAP.

[0207] Once the SAP reaction was complete, a primer extension reaction was initiated by adding a polymorphism-specific MassEXTEND™ primer cocktail to each sample. Each MassEXTEND™ cocktail included a specific combination of dideoxynucleotides (ddNTPs) and deoxynucleotides (dNTPs) used to distinguish polymorphic alleles from one another. Methods for verifying, allelotyping and genotyping SNPs are disclosed, for example, in U.S. Pat. No. 6,258,538, the content of which is hereby incorporated by reference. In Table 4, ddNTPs are shown and the fourth nucleotide not shown is the dNTP.

TABLE 4: Extension Primers

SNP Reference	Extend Probe	Termination Mix
rs132659	AGAACTCCCCAAATCGTCCT	ACG

[0208] The MassEXTEND™ reaction was performed in a total volume of 9 μl, with the addition of 1X ThermoSequenase buffer, 0.576 units of ThermoSequenase (Amersham Pharmacia), 600 nM MassEXTEND™ primer, 2 mM of ddATP and/or ddCTP and/or ddGTP and/or ddTTP, and 2 mM of dATP or dCTP or dGTP or dTTP. The deoxy nucleotide (dNTP) used in the assay normally was complementary to the nucleotide at the polymorphic site in the amplicon. Samples were incubated at 94°C for 2 minutes, followed by 55 cycles of 5 seconds at 94°C, 5 seconds at 52°C, and 5 seconds at 72°C.

[0209] Following incubation, samples were desalted by adding 16 μl of water (total reaction volume was 25 μl), 3 mg of SpectroCLEAN™ sample cleaning beads (Sequenom, Inc.) and allowed to incubate for 3 minutes with rotation. Samples were then robotically dispensed using a piezoelectric dispensing device (SpectroJET™ (Sequenom, Inc.)) onto either 96-spot or 384-spot silicon chips containing a matrix that crystallized each sample (SpectroCHIP™ (Sequenom, Inc.)). Subsequently, MALDI-TOF mass spectrometry (Biflex and Autoflex MALDI-TOF mass spectrometers (Bruker Daltonics) can be used) and

SpectroTYPER RT™ software (Sequenom, Inc.) were used to analyze and interpret the SNP genotype for each sample.

Genetic Analysis

[0210] Minor allelic frequencies for the polymorphisms set forth in Table A were verified as being 10% or greater using the extension assay described above in a group of samples isolated from 92 individuals originating from the state of Utah in the United States, Venezuela and France (Coriell cell repositories).

[0211] Genotyping results are shown for female pools in Table 5. In Table 5, "AF" refers to allelic frequency; and "F case" and "F control" refer to female case and female control groups, respectively.

TABLE 5: Genotyping Results

SNP Reference	AF F case	AF F control	p-value
rs132659	C = 0.675 T = 0.325	C = 0.589 T = 0.411	0.0015

[0212] The single marker alleles set forth in Table A were considered validated, since the genotyping data agreed with the allelotyping data and each SNP significantly associated with osteoarthritis. Particularly significant associations with osteoarthritis are indicated by a calculated p-value of less than 0.05 for genotype results.

Example 3

Association of Polymorphic Variants with Osteoarthritis in Replication Cohorts

[0213] The single marker polymorphism set forth in Table A was genotyped again in two replication cohorts consisting of individuals selected for OA.

Sample Selection and Pooling Strategies – Replication Sample 1

[0214] A second case control sample (replication sample #1) was created by using 100 Caucasian female cases from Chingford, UK, and 148 unrelated female cases from the St. Thomas twin study. Cases were defined as having Kellgren-Lawrence (KL) scores of at least 2 in at least one knee x-ray. In addition, 199 male knee replacement cases from Nottingham were included. (For a cohort description, see the Nottingham description provided in Example 1). The control pool was made up of unrelated female individuals from the St. Thomas twin study (England) with normal knee x-rays and without other indications of OA, regardless of anatomical location, as well as lacking family history of OA. The St. Thomas twin study consists of Caucasian, female participants from the St. Thomas' Hospital, London, adult-twin registry, which is a voluntary registry of >4,000 twin pairs ranging from 18 to 76 years of age.

The replication sample 1 cohort was used to replicate the initial results. Table 6 below summarizes the selected phenotype data collected from the case and control individuals.

TABLE 6

Phenotype	**Female cases (n=248):** median (range)/ (n,%)	**Male cases (n=199):** median (range)/ (n,%)	**Female controls (n=313):** mean (range)/ (n,%)
Age	59 (39- 73)	66 (45- 73)	55 (50- 72)
Height (cm)	162 (141- 178)	175 (152- 198)	162 (141- 176)
Weight (kg)	68 (51- 123)	86 (62- 127)	64 (40- 111)
Body mass index (kg/m^2)	26 (18- 44)	29 (21- 41)	24 (18- 46)
Kellgren-Lawrence* left knee	0 (63, 26%), 1 (20, 8%), 2 (105, 43%), 3 (58, 23%), 4 (1, 0%)	NA	NA
Kellgren-Lawrence* right knee	0 (43, 7%), 1 (18, 7%), 2 (127, 52%), 3 (57, 23%), 4 (1, 0%)	NA	NA
KL* >2 both knees	No (145, 59%), Yes (101, 41%)	NA	NA
KL* >2 either knee	No (0, 0%), Yes (248, 100%)	NA	NA

* 0: normal, 1: doubtful, 2: definite osteophyte (bony protuberance), 3: joint space narrowing (with or without osteophyte), 4: joint deformity

Sample Selection and Pooling Strategies – Replication Sample 2

[0215] A third case control sample (replication sample #2) was created by using individuals with symptoms of OA from Newfoundland, Canada. These individuals were recruited and examined by rheumatologists. Affected joints were x-rayed and a final diagnosis of definite or probable OA was made according to American College of Rheumatology criteria by a single rheumatologist to avoid any inter-examiner diagnosis variability. Controls were recruited from volunteers without any symptoms from the musculoskeletal system based on a normal joint exam performed by a rheumatologist. Only cases with a diagnosis of definite OA were included in the study. Only individuals of Caucasian origin were included. The cases consisted of 228 individuals with definite knee OA, 106 individuals with definite hip OA, and 74 individuals with hip OA.

TABLE 7

Phenotype	**Case**	**Control**
Age at Visit	62.7	52.5
Sex (Female/Male)	227/119	174/101
Knee OA Xray: No	35% (120)	80% (16)

Unknown	1% (4)	0% (0)
Yes	64% (221)	20% (4)
Hip OA Xray: No	63% (215)	80% (16)
Unknown	2% (7)	0% (0)
Yes	35% (121)	20% (4)

Assay for Verifying, Allelotyping, and Genotyping SNPs

[0216] Genotyping of the replication cohorts described in Tables 6 and 7 was performed using the same methods used for the original genotyping, as described herein. A MassARRAY™ system (Sequenom, Inc.) was utilized to perform SNP genotyping in a high-throughput fashion. This genotyping platform was complemented by a homogeneous, single-tube assay method (hME™ or homogeneous MassEXTEND™ (Sequenom, Inc.)) in which two genotyping primers anneal to and amplify a genomic target surrounding a polymorphic site of interest. A third primer (the MassEXTEND™ primer), which is complementary to the amplified target up to but not including the polymorphism, was then enzymatically extended one or a few bases through the polymorphic site and then terminated.

[0217] For each polymorphism, SpectroDESIGNER™ software (Sequenom, Inc.) was used to generate a set of PCR primers and a MassEXTEND™ primer which where used to genotype the polymorphism. Other primer design software could be used or one of ordinary skill in the art could manually design primers based on his or her knowledge of the relevant factors and considerations in designing such primers. Table 3 shows PCR primers and Table 4 shows extension probes used for analyzing (*e.g.*, genotyping) polymorphisms in the replication cohorts. The initial PCR amplification reaction was performed in a 5 µl total volume containing 1X PCR buffer with 1.5 mM $MgCl_2$ (Qiagen), 200 µM each of dATP, dGTP, dCTP, dTTP (Gibco-BRL), 2.5 ng of genomic DNA, 0.1 units of HotStar DNA polymerase (Qiagen), and 200 nM each of forward and reverse PCR primers specific for the polymorphic region of interest.

[0218] Samples were incubated at 95°C for 15 minutes, followed by 45 cycles of 95°C for 20 seconds, 56°C for 30 seconds, and 72°C for 1 minute, finishing with a 3 minute final extension at 72°C. Following amplification, shrimp alkaline phosphatase (SAP) (0.3 units in a 2 µl volume) (Amersham Pharmacia) was added to each reaction (total reaction volume was 7 µl) to remove any residual dNTPs that were not consumed in the PCR step. Samples were incubated for 20 minutes at 37°C, followed by 5 minutes at 85°C to denature the SAP.

[0219] Once the SAP reaction was complete, a primer extension reaction was initiated by adding a polymorphism-specific MassEXTEND™ primer cocktail to each sample. Each MassEXTEND™ cocktail included a specific combination of dideoxynucleotides (ddNTPs) and deoxynucleotides (dNTPs)

used to distinguish polymorphic alleles from one another. Methods for verifying, allelotyping and genotyping SNPs are disclosed, for example, in U.S. Pat. No. 6,258,538, the content of which is hereby incorporated by reference. In Table 7, ddNTPs are shown and the fourth nucleotide not shown is the dNTP.

[0220] The MassEXTEND™ reaction was performed in a total volume of 9 µl, with the addition of 1X ThermoSequenase buffer, 0.576 units of ThermoSequenase (Amersham Pharmacia), 600 nM MassEXTEND™ primer, 2 mM of ddATP and/or ddCTP and/or ddGTP and/or ddTTP, and 2 mM of dATP or dCTP or dGTP or dTTP. The deoxy nucleotide (dNTP) used in the assay normally was complementary to the nucleotide at the polymorphic site in the amplicon. Samples were incubated at 94°C for 2 minutes, followed by 55 cycles of 5 seconds at 94°C, 5 seconds at 52°C, and 5 seconds at 72°C.

[0221] Following incubation, samples were desalted by adding 16 µl of water (total reaction volume was 25 µl), 3 mg of SpectroCLEAN™ sample cleaning beads (Sequenom, Inc.) and allowed to incubate for 3 minutes with rotation. Samples were then robotically dispensed using a piezoelectric dispensing device (SpectroJET™ (Sequenom, Inc.)) onto either 96-spot or 384-spot silicon chips containing a matrix that crystallized each sample (SpectroCHIP™ (Sequenom, Inc.)). Subsequently, MALDI-TOF mass spectrometry (Biflex and Autoflex MALDI-TOF mass spectrometers (Bruker Daltonics) can be used) and SpectroTYPER RT™ software (Sequenom, Inc.) were used to analyze and interpret the SNP genotype for each sample.

<u>Genetic Analysis</u>

[0222] Genotyping results for replication cohorts #1 and #2 are provided in Tables 8 and 9, respectively.

TABLE 8

rsID	Replication #1 (Mixed Male/Female cases and Female controls)				Meta-analysis Disc. + Rep #1
	AF OA Con	AF OA Cas	Delta	P-value	P-value
rs132659	0.38	0.34	0.04	0.128	0.0077

TABLE 9

rsID	Replication #2 (Newfoundland) (Male/Female cases and controls)				Meta-analysis Disc. + Rep #2
	AF OA Con	AF OA Cas	Delta	P-value	Not Done
rs132659	0.36	0.36	-0.001	0.973	

[0223] To combine the evidence for association from multiple sample collections, a meta-analysis procedure was employed. The allele frequencies were compared between cases and controls within the discovery sample, as well as within the replication cohort #1 using the DerSimian-Laird approach (DerSimonian, R. and N. Laird. 1986. Meta-analysis in clinical trials. Control Clin Trials 7: 177-188.)

[0224] The absence of a statistically significant association in one or more of the replication cohorts should not be interpreted as minimizing the value of the original finding. There are many reasons why a biologically derived association identified in a sample from one population would not replicate in a sample from another population. The most important reason is differences in population history. Due to bottlenecks and founder effects, there may be common disease predisposing alleles present in one population that are relatively rare in another, leading to a lack of association in the candidate region. Also, because common diseases such as arthritis-related disorders are the result of susceptibilities in many genes and many environmental risk factors, differences in population-specific genetic and environmental backgrounds could mask the effects of a biologically relevant allele. For these and other reasons, statistically strong results in the original, discovery sample that did not replicate in one or more of the replication samples may be further evaluated in additional replication cohorts and experimental systems.

Example 4

APOL3 Region Proximal SNPs

[0225] It has been discovered that SNP rs132659 in *APOL3* is associated with occurrence of osteoarthritis in subjects. *APOL3* belongs to the high density lipoprotein family that plays a central role in cholesterol transport. The cholesterol content of membranes is important in cellular processes such as modulating gene transcription and signal transduction both in the adult brain and during neurodevelopment. It has been shown that the APOL1-APOL4 gene cluster on chromosome 22 exists only in primates (humans and African green monkeys) and not in dogs, pigs, or rodents, suggesting that this gene cluster has arisen recently in evolution (Monajemi et. al., *Genomics* 79: 539-546, 2002). Six transcript variants encoding three different isoforms have been identified.

[0226] Expression of this gene is upregulated by tumor necrosis factor-alpha in endothelial cells lining the normal and atherosclerotic iliac artery and aorta (Horrevoets et. al., *Blood* 93: 3418-3431, 1999). *APOL3* is genetically linked to OA and may play a role in the pathophysiology of OA brought about by inflammation. *APOL3* is likely inhibited by a small molecule inhibitor or by specific antibodies. *APOL3* activity may be increased in a subject by administering *APOL3* recombinant protein or a functional fragment thereof.

[0227] Two hundred-nineteen additional allelic variants proximal to rs132659 were identified and subsequently allelotyped in osteoarthritis case and control sample sets as described in Examples 1 and 2.

The polymorphic variants are set forth in Table 10. The chromosome positions provided in column four of Table 10 are based on Genome "Build 34" of NCBI's GenBank.

TABLE 10

dbSNP rs#	Chromosome	Position in SEQ ID NO: 1	Chromosome Position	Allele Variants
rs3888818	22	201	34781551	c/t
rs2010605	22	425	34781775	a/g
rs743919	22	1095	34782445	g/t
rs1008134	22	2201	34783551	a/c
rs132607	22	7879	34789229	a/g
rs1476029	22	8395	34789745	c/t
rs1476030	22	8461	34789811	c/t
rs2413380	22	9503	34790853	c/t
rs2051609	22	10304	34791654	g/t
rs2413381	22	10695	34792045	c/t
rs1894604	22	16300	34797650	a/g
rs1894605	22	16444	34797794	g/t
rs132609	22	17591	34798941	c/t
rs132610	22	17988	34799338	-/a
rs132611	22	19116	34800466	-/t
rs132612	22	19358	34800708	c/t
rs1008790	22	20300	34801650	a/g
rs23085	22	20669	34802019	a/t
rs105161	22	20891	34802241	a/g
rs132613	22	21451	34802801	c/t
rs132614	22	21978	34803328	c/t
rs132615	22	22785	34804135	c/g
rs132617	22	24248	34805598	c/t
rs3865724	22	24770	34806120	c/t
rs2019657	22	24844	34806194	a/g
rs3865725	22	25066	34806416	g/t
rs2019364	22	25096	34806446	c/t
rs2008383	22	25309	34806659	a/g
rs3986002	22	25344	34806694	a/c
rs3888942	22	25529	34806879	a/t
rs3888943	22	25537	34806887	a/g
rs3888944	22	25554	34806904	a/c
rs132618	22	27963	34809313	a/t
rs132619	22	28134	34809484	g/t
rs3827346	22	28356	34809706	a/g
rs132620	22	29648	34810998	-/a
rs132621	22	29986	34811336	a/g
rs80575	22	30217	34811567	g/t
rs80576	22	30267	34811617	a/g
rs80577	22	30315	34811665	a/g
rs80578	22	30585	34811935	c/t
rs80579	22	30724	34812074	a/c
rs80580	22	30897	34812247	c/t

dbSNP rs#	Chromosome	Position in SEQ ID NO: 1	Chromosome Position	Allele Variants
rs132622	22	30931	34812281	c/t
rs132623	22	31080	34812430	g/t
rs132624	22	31246	34812596	c/t
rs132625	22	31373	34812723	a/t
rs132626	22	31463	34812813	a/g
rs132627	22	31467	34812817	a/g
rs1807672	22	32188	34813538	g/t
rs132628	22	32288	34813638	c/t
rs132629	22	32520	34813870	a/t
rs132630	22	32594	34813944	a/c
rs132631	22	32657	34814007	a/c
rs132632	22	32677	34814027	a/g
rs132633	22	32764	34814114	c/t
rs132634	22	32784	34814134	a/g
rs132635	22	32830	34814180	c/t
rs132636	22	32872	34814222	c/t
rs129603	22	33121	34814471	a/c
rs132637	22	33348	34814698	g/t
rs3788518	22	33952	34815302	c/g
rs132638	22	34184	34815534	c/g
rs132639	22	34361	34815711	a/t
rs132640	22	35026	34816376	a/g
rs132641	22	35192	34816542	a/g
rs132642	22	35600	34816950	a/t
rs132643	22	36033	34817383	c/t
rs132644	22	36289	34817639	c/t
rs132645	22	38869	34820219	a/g
rs2017329	22	39629	34820979	a/t
rs739198	22	40530	34821880	c/t
rs132647	22	41621	34822971	c/t
rs2097465	22	42379	34823729	c/t
rs2105915	22	42802	34824152	c/t
rs132648	22	42865	34824215	t/c
rs132649	22	43644	34824994	a/g
rs132650	22	45051	34826401	c/t
rs132651	22	45828	34827178	a/c
rs132652	22	45829	34827179	a/t
rs80584	22	46257	34827607	c/t
rs132653	22	47286	34828636	a/c
rs916334	22	47427	34828777	c/t
rs132654	22	47963	34829313	c/t
rs132655	22	48013	34829363	c/t
rs132656	22	48229	34829579	c/t
rs3834684	22	48282	34829632	-/a
rs132657	22	48376	34829726	-/g
rs916335	22	48404	34829754	a/g
rs132659	22	49900	34831250	c/t

dbSNP rs#	Chromosome	Position in SEQ ID NO: 1	Chromosome Position	Allele Variants
rs132660	22	52699	34834049	g/t
rs132661	22	52897	34834247	a/g
rs132662	22	53414	34834764	a/g
rs132663	22	53487	34834837	a/t
rs132664	22	54112	34835462	g/t
rs132667	22	55492	34836842	a/g
rs132670	22	59766	34841116	c/t
rs132671	22	60307	34841657	a/g
rs132672	22	60701	34842051	a/g
rs132673	22	60952	34842302	a/g
rs132674	22	61401	34842751	c/t
rs132675	22	62379	34843729	c/t
rs80585	22	62870	34844220	c/t
rs80586	22	62879	34844229	a/g
rs132676	22	63499	34844849	a/t
rs132677	22	64284	34845634	-/a
rs132678	22	64408	34845758	a/g
rs132680	22	64760	34846110	a/g
rs132681	22	65230	34846580	a/g
rs132683	22	66127	34847477	a/g
rs2269594	22	66634	34847984	c/t
rs132684	22	66686	34848036	a/g
rs132685	22	66694	34848044	c/g
rs132686	22	67113	34848463	a/g
rs132687	22	67257	34848607	a/g
rs132688	22	67403	34848753	a/g
rs132689	22	67609	34848959	a/g
rs132690	22	68418	34849768	-/a
rs132691	22	68610	34849960	c/g
rs132692	22	69629	34850979	c/t
rs132693	22	70024	34851374	a/g
rs132694	22	70848	34852198	a/g
rs132695	22	71428	34852778	c/g
rs1966266	22	71553	34852903	c/t
rs1966267	22	71633	34852983	a/g
rs106808	22	71768	34853118	a/c
rs132696	22	71769	34853119	a/g
rs2239829	22	73039	34854389	a/g
rs2285154	22	73325	34854675	a/g
rs2239830	22	73412	34854762	a/c
rs2239831	22	73547	34854897	c/t
rs3865722	22	73769	34855119	a/t
rs3865723	22	73806	34855156	a/g
rs3985996	22	74467	34855817	c/t
rs3985997	22	74472	34855822	c/t
rs3985998	22	74473	34855823	a/g
rs3985999	22	74482	34855832	c/t

dbSNP rs#	Chromosome	Position in SEQ ID NO: 1	Chromosome Position	Allele Variants
rs3986000	22	74494	34855844	a/c
rs2413382	22	74592	34855942	a/g
rs2413383	22	74670	34856020	g/t
rs2413384	22	74672	34856022	g/t
rs2413385	22	74714	34856064	g/t
rs2413386	22	74723	34856073	a/t
rs1894606	22	74749	34856099	a/g
rs916336	22	74861	34856211	c/g
rs916337	22	74892	34856242	c/t
rs916338	22	74893	34856243	c/t
rs132697	22	75176	34856526	a/g
rs12781	22	75705	34857055	a/g
rs1053983	22	75989	34857339	a/g
rs1053982	22	76027	34857377	a/g
rs2227167	22	77949	34859299	a/g
rs2227168	22	77974	34859324	c/t
rs132700	22	78167	34859517	c/t
rs3075364	22	78310	34859660	-/ct
rs2227169	22	78415	34859765	c/t
rs2097466	22	78575	34859925	c/t
rs2097467	22	78590	34859940	c/t
rs2413387	22	78709	34860059	c/t
rs132701	22	78875	34860225	c/t
rs132702	22	79864	34861214	c/t
rs132703	22	81316	34862666	c/t
rs2269595	22	81320	34862670	a/g
rs2269596	22	81409	34862759	c/t
rs132704	22	81737	34863087	c/t
rs2007468	22	81843	34863193	a/g
rs132705	22	82102	34863452	c/t
rs2007706	22	82833	34864183	c/t
rs132706	22	83461	34864811	c/t
rs132707	22	83624	34864974	c/t
rs132708	22	83660	34865010	c/g
rs132709	22	83701	34865051	a/t
rs132710	22	83708	34865058	a/g
rs132711	22	83782	34865132	c/t
rs132712	22	85707	34867057	a/g
rs132713	22	85717	34867067	a/g
rs132714	22	86486	34867836	c/t
rs132716	22	86833	34868183	a/g
rs132717	22	87115	34868465	c/t
rs132718	22	87234	34868584	a/g
rs132719	22	87479	34868829	g/t
rs132720	22	87561	34868911	a/g
rs132721	22	87604	34868954	a/g
rs132722	22	87674	34869024	c/t

dbSNP rs#	Chromosome	Position in SEQ ID NO: 1	Chromosome Position	Allele Variants
rs132723	22	87958	34869308	a/g
rs132724	22	87992	34869342	-/g
rs132725	22	88019	34869369	a/g
rs132726	22	88074	34869424	c/g
rs132727	22	88079	34869429	c/g
rs132728	22	88115	34869465	a/g
rs132729	22	88118	34869468	c/g
rs132730	22	88120	34869470	a/g
rs132731	22	88135	34869485	-/ctcat
rs132732	22	88142	34869492	g/t
rs132733	22	88143	34869493	g/t
rs140575	22	88149	34869499	aca/tg
rs132734	22	88340	34869690	a/g
rs132735	22	88344	34869694	g/t
rs80587	22	88512	34869862	c/g
rs132736	22	88521	34869871	c/t
rs132737	22	88650	34870000	c/g
rs132738	22	88827	34870177	c/t
rs1807673	22	89230	34870580	a/g
rs2014700	22	89236	34870586	a/g
rs132739	22	90754	34872104	g/a
rs1812023	22	90984	34872334	a/g
rs1812024	22	91110	34872460	a/g
rs2005590	22	92026	34873376	c/t
rs132740	22	92954	34874304	c/t
rs3986001	22	93375	34874725	-/ttgc
rs2413390	22	93794	34875144	c/t
rs132743	22	94937	34876287	c/g
rs132744	22	95068	34876418	c/t
rs2413391	22	96188	34877538	a/g
rs132749	22	97092	34878442	c/t
rs132750	22	98812	34880162	c/t
rs132741	22	not mapped	not mapped	a/c
rs2413388	22	not mapped	not mapped	a/t
rs2413389	22	not mapped	not mapped	c/g

Assay for Verifying and Allelotyping SNPs

[0228] The methods used to verify and allelotype the 219 proximal SNPs of Table 10 are the same methods described in Examples 1 and 2 herein. The primers and probes used in these assays are provided in Table 11 and Table 12, respectively.

TABLE 11

dbSNP rs#	Forward PCR primer	Reverse PCR primer

dbSNP rs#	Forward PCR primer	Reverse PCR primer
rs3888818	ACGTTGGATGTGAGGTCAGGAGTTTGAGAC	ACGTTGGATGCCATGCCCAGCTAATTTTCG
rs2010605	ACGTTGGATGTGTGCTTTTTATGTCTTAGG	ACGTTGGATGGACTTTTAGAAGAAAAGTAC
rs743919	ACGTTGGATGTTCTTCACCAAGCCCTCTTC	ACGTTGGATGCCCAACACACACAAAGATGG
rs1008134	ACGTTGGATGACATATCCGGGCATCTTTTC	ACGTTGGATGCATCCACAGGATGCAATATC
rs132607	ACGTTGGATGTAGTTTGCAGGTCACAAGGG	ACGTTGGATGGGAAGGAAGACCCAAACAGC
rs1476029	ACGTTGGATGTGGGTGACAGAGATCCTTAC	ACGTTGGATGAAACTCAGGGAAACGGACTC
rs1476030	ACGTTGGATGATCCTAGCACTGGGATTTGG	ACGTTGGATGCGTTTCCCTGAGTTTCACTG
rs2413380	ACGTTGGATGATGATACTGAGTCCAGGAGG	ACGTTGGATGAAAGGCTACTTCTTGCTCAC
rs2051609	ACGTTGGATGTAATCCCAGCACTTTGGGAG	ACGTTGGATGACAGACGGGGTTTCATCATG
rs2413381	ACGTTGGATGAGGGCTGCAGTAGAAAAGCG	ACGTTGGATGACGATGCGTGTGCCGACAG
rs1894604	ACGTTGGATGAAGTGCTGCTGCAAAAAGAG	ACGTTGGATGTTCTCCACTTCCATTCTGTG
rs1894605	ACGTTGGATGTGATGAGAGATGCAGATCGC	ACGTTGGATGTATTCAGAATTCAGCCTGCG
rs132609	ACGTTGGATGGCATTTGAAAGGTCCGTATC	ACGTTGGATGCCCAAATCTGTCTTTTAGCC
rs132610	ACGTTGGATGTTCTACAAGAGCTAGGGACC	ACGTTGGATGGATCTATTGCTGCTTAGGCC
rs132611	ACGTTGGATGGCCTTCTTTACTCTGTCCTC	ACGTTGGATGGGTTTTCTTTCAGGTCCTCC
rs132612	ACGTTGGATGAAAAATCTTCCCGCTACCTC	ACGTTGGATGTTCTGTGAGCTTTCTCTCTG
rs1008790	ACGTTGGATGGTGAGGTCCCTTTTATGATG	ACGTTGGATGATAACAGCCCCTGACAGATG
rs23085	ACGTTGGATGAAACCGTGCCAGCTGAGGAT	ACGTTGGATGGTCGGCAAGGAAGAGGAATC
rs105161	ACGTTGGATGAAGGAGCGGGAAATCTTTTG	ACGTTGGATGGTAGGAGGCTGAAATGCTAG
rs132613	ACGTTGGATGATGTAAAACCAATGGCCTCC	ACGTTGGATGAAGCTTCAGATTGTTCACCC
rs132614	ACGTTGGATGCAAGAGCCCTGCTTTGTGAG	ACGTTGGATGTCCTGCACCAGCAGAGATGA
rs132615	ACGTTGGATGAAATCTGGAGGCTTGGTGAC	ACGTTGGATGTGAGCATTCACATGGGACAG
rs132617	ACGTTGGATGGAGAAGAAGAGTGTGTGTGC	ACGTTGGATGACAGCCACCTGAATTTGTGC
rs3865724	ACGTTGGATGTTCCTGGATAATTCCCATTC	ACGTTGGATGGCTGGATCACTGAAGAAGTA
rs2019657	ACGTTGGATGACGCCAGAACATTGTGTTTC	ACGTTGGATGGTGCCAGAAACATTCAAAGC
rs3865725	ACGTTGGATGAATATAGAACTGCTGGGCGC	ACGTTGGATGTGACTTAGGAGAGGTTCTGG
rs2019364	ACGTTGGATGTGACTTAGGAGAGGTTCTGG	ACGTTGGATGAATATAGAACTGCTGGGCGC
rs2008383	ACGTTGGATGTGATTCTAGGAGCAGGACTG	ACGTTGGATGACATGGGTGACCCTATCAAG
rs3986002	ACGTTGGATGCTTCTGTCCTTCTCTGTGTC	ACGTTGGATGCAGGCAGAGGATTTGTTTGG
rs3888942	ACGTTGGATGCTGGGCTTTTGTGCTAAGAG	ACGTTGGATGGGGCCAATTTGCCCCATAAA
rs3888943	ACGTTGGATGCTGGGCTTTTGTGCTAAGAG	ACGTTGGATGTTGGGCCAATTTGCCCCATA
rs3888944	ACGTTGGATGATACAGCCCTTGCCACTATG	ACGTTGGATGTTGAAGACATGGAAGCAGGG
rs132618	ACGTTGGATGAATCCGTGCCATCAGGCAAG	ACGTTGGATGCCTGCAATCGTTCTCTCTGC
rs132619	ACGTTGGATGTCATCAGCAGAAGCTGAAGC	ACGTTGGATGGGTGTGATGTCACGCATAAC
rs3827346	ACGTTGGATGAAGAGGTCCACAGAGGCTG	ACGTTGGATGAAAACAAGACCAGCAAGGGA
rs132620	ACGTTGGATGTCACATTAGATCAGGAAGCC	ACGTTGGATGTTAGGCCAGTTTAGCAGAAA
rs132621	ACGTTGGATGCTTCAAATCTGCAACTGGTG	ACGTTGGATGGATAGCTTAAGGACTCAGAG
rs80575	ACGTTGGATGGCTGCACATGAACTCTCAAG	ACGTTGGATGTGACATGTGACAGTGAGACC
rs80576	ACGTTGGATGTGAAGCTGTCACCTGCTAAG	ACGTTGGATGAACTCTCAAGCCACTTGACG
rs80577	ACGTTGGATGAGTCCATAAGAGGTTCCATG	ACGTTGGATGAACTAATGCCTTAGCAGGTG
rs80578	ACGTTGGATGACTGTTTCCCTGACAGCATG	ACGTTGGATGTGTAGAACAGAAGAGGGTCC
rs80579	ACGTTGGATGTGGGAGTAGGGTGAGAAGAG	ACGTTGGATGACTCACTGGTCCTCTGCAAG
rs80580	ACGTTGGATGTTCAATCAGATGGGCGTGTG	ACGTTGGATGGATGGCATCATGCTACTTGG
rs132622	ACGTTGGATGTATGTCTTGGAGACTGGGAC	ACGTTGGATGACCTGCTGTTCATTCTCAGG
rs132623	ACGTTGGATGAGCTCTGTCCAACTCCATTC	ACGTTGGATGCTGAGGAACTGCACAAACAC
rs132624	ACGTTGGATGTGCTGGGATTACAGGCATGA	ACGTTGGATGTCAAAGAAAGTCCTGCTGGG
rs132625	ACGTTGGATGTTTCACGCCATTCTCCTGCC	ACGTTGGATGCGATGAAACCCCGTCTGTAC

dbSNP rs#	Forward PCR primer	Reverse PCR primer
rs132626	ACGTTGGATGTGGAGTGCAGTGGCATGATC	ACGTTGGATGGCAGGAGAATGGCGTGAAAC
rs132627	ACGTTGGATGAGGAGAATGGCGTGAAACCG	ACGTTGGATGAGACAGAGTCTTGCTTGTCC
rs1807672	ACGTTGGATGGTGTGCTACAGCCTAAATGG	ACGTTGGATGAATACCCCATGTGACAGCTG
rs132628	ACGTTGGATGTATAGACTGAGTTGTGTGCC	ACGTTGGATGTCCTTAAAGGCTCAATCTCC
rs132629	ACGTTGGATGCTCTCTCCCTGTCTCTCTTT	ACGTTGGATGTGTGTCCTCACATGGCCTTC
rs132630	ACGTTGGATGTTCCAAGGTGAAGGTGCCAG	ACGTTGGATGAAGGCCATGTGAGGACACAG
rs132631	ACGTTGGATGGGTGGCTCCAACAACTGATG	ACGTTGGATGATCAACCCTGCTGGCACCTT
rs132632	ACGTTGGATGCTTGGAATTTTTGCCTCCAG	ACGTTGGATGTCAGGATGCCTTAGTAAAAC
rs132633	ACGTTGGATGAGAAGAGTGATTCACCAGGG	ACGTTGGATGGGAAAGCTCACTTTCTGGTG
rs132634	ACGTTGGATGAAGTGCCATGGTGCTTTGTG	ACGTTGGATGGAAAGCATGGTGGAAAGCTC
rs132635	ACGTTGGATGAATAGGCACATGGCAGAAGG	ACGTTGGATGCACCAGAAAGTGAGCTTTCC
rs132636	ACGTTGGATGAAGCGTTTGACAATAGGCAC	ACGTTGGATGAAAGTGAGCTTTCCACCATG
rs129603	ACGTTGGATGGTGTCATATTGACACAGATTG	ACGTTGGATGAGGGTGTATATATATATACCC
rs132637	ACGTTGGATGGCATCTTAGTACACAGCAGG	ACGTTGGATGTTCCCAAATCCCTGCAAACC
rs3788518	ACGTTGGATGAATCCTTCAGAAGGGCTTGG	ACGTTGGATGGCCGCGTTATTAAACCACAG
rs132638	ACGTTGGATGCATCCTTTCAGTGAAGGAGG	ACGTTGGATGTTGCCAAGGCAACTCAGTGA
rs132639	ACGTTGGATGACACCTGGGCAAACAAAAGC	ACGTTGGATGAAGTTCCCCATAGTTGGCAG
rs132640	ACGTTGGATGTAAGAAGCTCCAGGTGACAC	ACGTTGGATGAAAAGAGTGACTCAGCGTCC
rs132641	ACGTTGGATGAGGGTCAGCTGGGAGCAGA	ACGTTGGATGAGGGCTGAGAGAGGAGGTTG
rs132642	ACGTTGGATGAAGAAGCAAGCCTACCTGAG	ACGTTGGATGAAACGAACCCTTCCAGTCAG
rs132643	ACGTTGGATGATCACAGACACCCAGTACAC	ACGTTGGATGACGTTCTGACAATGACCTGG
rs132644	ACGTTGGATGGCATAGAGTGCAAGACACAG	ACGTTGGATGGGGCTCCACTCCCTTAAATA
rs132645	ACGTTGGATGTGAAGGCAAACAGTACAAGA	ACGTTGGATGAAGTTAACCAAGTGTTTAC
rs2017329	ACGTTGGATGCCTTCCCAATTAAAAGCAGC	ACGTTGGATGGGGCAACAAGAGTGAAATTC
rs739198	ACGTTGGATGAAACTTTGGTCTCCACAACC	ACGTTGGATGTGAGTTTGTCTAAAGACCGG
rs132647	ACGTTGGATGCCTCACTACAGAAACCATGG	ACGTTGGATGAACTCAACTGGTTCAACCAC
rs2097465	ACGTTGGATGGAATTGACCAAACTGCAGGC	ACGTTGGATGAGGGTTGAAGCTGGATACTG
rs2105915	ACGTTGGATGAACCCAGGAGTTCAGGACAA	ACGTTGGATGGGGAACTACAAGTGCATCAC
rs132648	ACGTTGGATGGTGGCTCAGGGCTGTAATTC	ACGTTGGATGTGTCCTGAACTCCTGGGTTC
rs132649	ACGTTGGATGGTCCTCCCCAGTCTTATTAC	ACGTTGGATGATTCAGAGGTTAGCTGGCTC
rs132650	ACGTTGGATGAAAGTGCTGGGATTACAGGC	ACGTTGGATGCTAAATCTCCTGCCATAGGG
rs132651	ACGTTGGATGAGGTCAGGTGTTGACCTTCC	ACGTTGGATGAGCAGGGTAGGGCATCCTAA
rs132652	ACGTTGGATGAGCAGGGTAGGGCATCCTAAC	ACGTTGGATGAGGTCAGGTGTTGACCTTCC
rs80584	ACGTTGGATGCATGGAGTCCTGTGATCTAC	ACGTTGGATGAAACTGAGGCCATGGGAGAT
rs132653	ACGTTGGATGCAGCCGTGCATCTGCATAAT	ACGTTGGATGCACTTTCCCTTTTGGGTTCC
rs916334	ACGTTGGATGAAACAGGATGCTTCCCAGCC	ACGTTGGATGCTGCTCTTGGATCAGCAGGA
rs132654	ACGTTGGATGTAAGGAAGTGTCCAGAAGCC	ACGTTGGATGAAATGATCCTCCTGCCTCAG
rs132655	ACGTTGGATGACTTTTCCAGGTGAAGGTAG	ACGTTGGATGTTGTCGAACGCCATACCTGT
rs132656	ACGTTGGATGTCACTAGAAGCAAGGAACCC	ACGTTGGATGCCCAGGTTGACTGAACAAAG
rs3834684	ACGTTGGATGAGCCCTTGTTCACTAGAAGC	ACGTTGGATGGGTGGATGTGGGAGTAAAAG
rs132657	ACGTTGGATGGGCTGACTGACAATTACCTG	ACGTTGGATGAGGGCTCTGAGCTTTTCAAG
rs916335	ACGTTGGATGGATGACGAGAAAAAGGTGGG	ACGTTGGATGATTTGAAGGATGCAGTCTTG
rs132659	ACGTTGGATGGGCCCATAGTGGGTCATAAC	ACGTTGGATGGTGGGGTGAGTGCCCAAAAG
rs132660	ACGTTGGATGTACATGTGGTTGTACCCTCC	ACGTTGGATGCTGGCATGGTTTTACCCATC
rs132661	ACGTTGGATGCTTCGCAGAAATCATTCCGC	ACGTTGGATGCCAAAATGCAAGCTCAAGGC
rs132662	ACGTTGGATGCATCTCTTAAATGGGCCAGG	ACGTTGGATGTTGAAAGCCACAGCCTCATG
rs132663	ACGTTGGATGCCACTAAACGGATTGAGATC	ACGTTGGATGTGGCTTTCAACCAGCAACTC

dbSNP rs#	Forward PCR primer	Reverse PCR primer
rs132664	ACGTTGGATGATCATGCCACTGCAATCCAG	ACGTTGGATGGCATATGTGACTGCTTCCTC
rs132667	ACGTTGGATGCTGGAGAAATCAAATAGAGAG	ACGTTGGATGTGTACAGCTTTTGACAGTTG
rs132670	ACGTTGGATGTAAGGTCGGGAGTTCAAGAC	ACGTTGGATGACGCCCGGCTGATTTTGTAT
rs132671	ACGTTGGATGGTGAGCCATACCATCACATC	ACGTTGGATGCTGTAGTAAAGGTCTGGTCG
rs132672	ACGTTGGATGCTCCCCAATAAGCTCAACAC	ACGTTGGATGCTGTTAGGGCAATGAAAGGC
rs132673	ACGTTGGATGTGAGTAGTTGGTGTGAGTGG	ACGTTGGATGCAATGGATGAAGCTGATCCC
rs132674	ACGTTGGATGAACTGTAGTCCCAGCTACTC	ACGTTGGATGTAGCTCTATCACTCATGCTG
rs132675	ACGTTGGATGGTAGAGCAGATGTGCAATGG	ACGTTGGATGTCCTAACCATCTGCCTTGTG
rs80585	ACGTTGGATGCTGTTGTTCCAACACTTCAC	ACGTTGGATGGGTCTGCTACTAGAATTCAG
rs80586	ACGTTGGATGGTAAGTGTAAGAAGGTCTGC	ACGTTGGATGCAAGGCATAATATTCTGACC
rs132676	ACGTTGGATGCAAACATTCTGCAGAAAGCG	ACGTTGGATGAAGCGTGTTGCTGAGAAATG
rs132677	ACGTTGGATGCTCTGTTACAAAATGAAGGG	ACGTTGGATGGCTATCTAGGCTAAAAATCCC
rs132678	ACGTTGGATGAAGGCACTGAAAATGCCTAG	ACGTTGGATGGGAATCCAGATGCTTACATG
rs132680	ACGTTGGATGGCCTTAGCTATCATGTTCTC	ACGTTGGATGGCGTGTTTAAGGCAATTCTC
rs132681	ACGTTGGATGTACTGAAGCCTGAGACTAGC	ACGTTGGATGCTAGCAGAAACTAACCGAGC
rs132683	ACGTTGGATGTTACCCTATGGTAATGGCAG	ACGTTGGATGACTGATTAGTACAGGAAGGG
rs2269594	ACGTTGGATGTGGCATGGCTAAAAGGACAG	ACGTTGGATGGATTGTTCTGATGCCCAGTG
rs132684	ACGTTGGATGCCTTTTAGCCATGCCATTCG	ACGTTGGATGTCAGTGTAAAACGTGCCACC
rs132685	ACGTTGGATGTCAGTGTAAAACGTGCCACC	ACGTTGGATGCCTTTTAGCCATGCCATTCG
rs132686	ACGTTGGATGCAGAATATCCACGTCAGGTG	ACGTTGGATGGACAGCTTAGGACTATGTGC
rs132687	ACGTTGGATGCAACTGTAAGCAGCCCATTG	ACGTTGGATGCTGACGGTGCAAATGGATAC
rs132688	ACGTTGGATGAGTACTACAGGACGTGCTTG	ACGTTGGATGGGTCGCCTCATATATGGTAG
rs132689	ACGTTGGATGTACTGGGACAGTCTGCTTTC	ACGTTGGATGACTTTACAGTGCTGGAGCAG
rs132690	ACGTTGGATGTGTTTTGCTTTGCGCTCTCC	ACGTTGGATGTCTGCAACCAACTCTTTGGG
rs132691	ACGTTGGATGGTCAAAGCCAGGCATTTGTC	ACGTTGGATGCTGTCATCTTGTGGAAAGGG
rs132692	ACGTTGGATGGAATCTAAGCCAGCTGTTGG	ACGTTGGATGGGAGCATCATGTGGATTCCT
rs132693	ACGTTGGATGGCCAGAAGAAAAGAGTGTGG	ACGTTGGATGATTCTGCATGTGGAACGTCC
rs132694	ACGTTGGATGATAGAGACTGAGAGCTGCAG	ACGTTGGATGCAGAACAAAGCAGGAAGCTC
rs132695	ACGTTGGATGGCCTCTCTCTATGACTACAC	ACGTTGGATGTTCACAGCAGGGAACTCTTC
rs1966266	ACGTTGGATGTGATTGTACAAGGCAGACCC	ACGTTGGATGTGTAAGCACCTGCATTCAGC
rs1966267	ACGTTGGATGTAACTCACAGACCATGAGGG	ACGTTGGATGGGAGGAAAGCACAGCAGAAT
rs106808	ACGTTGGATGAACAAGGCAGATCCTTCCCG	ACGTTGGATGATGGTTCCTGAAGAGCAGTG
rs132696	ACGTTGGATGATGGTTCCTGAAGAGCAGTG	ACGTTGGATGAACAAGGCAGATCCTTCCCG
rs2239829	ACGTTGGATGTGTCTTTGTCGTTCGGATGG	ACGTTGGATGAAAGAGCGAAACTCCGTCTC
rs2285154	ACGTTGGATGTGAACTCAAATGATCCGCCC	ACGTTGGATGAAGAACCCTTTTCGACTGGG
rs2239830	ACGTTGGATGAAACCCTAATGGGAAGCCTC	ACGTTGGATGTGTGGTAGCAAGCAGTTGAC
rs2239831	ACGTTGGATGAGAACAGTCACTGACCCAAG	ACGTTGGATGGCTCCACACACTTTGATTCC
rs3865722	ACGTTGGATGCCACTGTACTGCTAGTATTG	ACGTTGGATGACCTGCTCTAGTTTTCATCC
rs3865723	ACGTTGGATGCCTGCATTTGATGCAATTCC	ACGTTGGATGGTTTCTGTTTCTGTTGCTTGC
rs3985996	ACGTTGGATGACATGGGTGACCCTATCAAG	ACGTTGGATGTGATTCTAGGAGCAGGACTG
rs3985997	ACGTTGGATGCAAGAATTTCTCCCGGCATC	ACGTTGGATGTGATTCTAGGAGCAGGACTG
rs3985998	ACGTTGGATGTGATTCTAGGAGCAGGACTG	ACGTTGGATGCAAGAATTTCTCCCGGCATC
rs3985999	ACGTTGGATGTGATTCTAGGAGCAGGACTG	ACGTTGGATGCAAGAATTTCTCCCGGCATC
rs3986000	ACGTTGGATGCAGGCAGAGGATTTGTTTGG	ACGTTGGATGCTTCTGTCCTTCTCTGTGTC
rs2413382	ACGTTGGATGAAACAAATCCTCTGCCTGGG	ACGTTGGATGAAAAGCCCAGAGCCTTCATG
rs2413383	ACGTTGGATGCTGGGCTTTTGTGCTAAGAG	ACGTTGGATGGGGCCAAGTTGACCCATAAA
rs2413384	ACGTTGGATGCTGGGCTTTTGTGCTAAGAG	ACGTTGGATGGGGCCAAGTTGACCCATAAA

dbSNP rs#	Forward PCR primer	Reverse PCR primer
rs2413385	ACGTTGGATGAATGGTCTTCGCTGATACAC	ACGTTGGATGATGGAAGCCGGTGTTTGATG
rs2413386	ACGTTGGATGTTTTATGGGTCAACTTGGCC	ACGTTGGATGTTCCAAAATGGTCTTCGCTG
rs1894606	ACGTTGGATGTTTTATGGGTCAACTTGGCC	ACGTTGGATGAGACTCCTGCAAAAGCTTCC
rs916336	ACGTTGGATGGGATGAGGGTATTTGCTGTC	ACGTTGGATGGGCCTGTATGTAGGTTGAAG
rs916337	ACGTTGGATGAGATCAGAAGGGCCTGTATG	ACGTTGGATGATTTGCTGTCTGGCTGTCTC
rs916338	ACGTTGGATGATTTGCTGTCTGGCTGTCTC	ACGTTGGATGAGATCAGAAGGGCCTGTATG
rs132697	ACGTTGGATGCATGAGGAAGAGAAGTCAGG	ACGTTGGATGACACTGACTGATGACTGAGC
rs12781	ACGTTGGATGTTACACACAGGGCACTCAGC	ACGTTGGATGGAGCCAGAAAATTAACTGAAAAGC
rs1053983	ACGTTGGATGCCATCATCAAGAAGCCACTG	ACGTTGGATGTGTCTTACCAGCATCCACTC
rs1053982	ACGTTGGATGGGAGAGTGGATGCTGGTAAG	ACGTTGGATGCGGTTGAATGTCCTTCCAAG
rs2227167	ACGTTGGATGCTCCTGAGTGTATGGACATC	ACGTTGGATGGGCATTAAGGGACATTCTGC
rs2227168	ACGTTGGATGCCAATTGGAGGCATTAAGGG	ACGTTGGATGCTCCTGAGTGTATGGACATC
rs132700	ACGTTGGATGAATGTGGTGTCTGGCTCCAC	ACGTTGGATGCTCAGCCCTGCTGTAAATGG
rs3075364	ACGTTGGATGGAACAGCAGTTTAGGGAGTG	ACGTTGGATGCGAAGCCTTTCTATGGACTC
rs2227169	ACGTTGGATGAGGTAAGTAAGCTGCCTTTC	ACGTTGGATGTTCAGAGCTTCATAGAGAGC
rs2097466	ACGTTGGATGCTGGGATTACAAGCATGAGC	ACGTTGGATGCTGCATAAATCACAGAGCTG
rs2097467	ACGTTGGATGATCTCCTGACCTTGTGATCC	ACGTTGGATGATTCTTTTCAAGGCCGGGCG
rs2413387	ACGTTGGATGAAGTAGCTGGGACTACAGGC	ACGTTGGATGTAACACGGTGAAACCCCGTC
rs132701	ACGTTGGATGGTGGCATATCTATGTTGTAC	ACGTTGGATGGCGAGACTCCATCTCAAAAA
rs132702	ACGTTGGATGGAAGCTCACCCAGTTAAGGA	ACGTTGGATGCCCCTGTAACAACAATCCTG
rs132703	ACGTTGGATGCTTGACCTGATCAATGTGTG	ACGTTGGATGTTTGTGCAGTTCCTCAGAAG
rs2269595	ACGTTGGATGAGAAGTTCAGGAAAAGGGCC	ACGTTGGATGCAGCAGGACTTTCTTTGGGA
rs2269596	ACGTTGGATGAGGTGCTCAGTTAGCGTTAC	ACGTTGGATGTCCCAAAGAAAGTCCTGCTG
rs132704	ACGTTGGATGATATTCTTCCTGCACTGCTG	ACGTTGGATGATCTCCCCGGGCTAGTTTTC
rs2007468	ACGTTGGATGAGGTTACCTGGGCAATTCAG	ACGTTGGATGGAAAATCCTGCTGACTAGCG
rs132705	ACGTTGGATGTTTTGATGGAGGCACCAGTG	ACGTTGGATGTCTCCAAATACGGTCACTGG
rs2007706	ACGTTGGATGCCCAGGAATTTACATAAGGG	ACGTTGGATGTTGAACATAGCAAGAGTGAG
rs132706	ACGTTGGATGAAGGATCAGTGCTGAGGGTC	ACGTTGGATGATTCCTCCTGCTGGTCATGG
rs132707	ACGTTGGATGAATCCTTAGGAAGGGCTGGG	ACGTTGGATGAGCTGGCCCCGTTAGTAAAC
rs132708	ACGTTGGATGTCTTGTTTCAGAGGGAGAGC	ACGTTGGATGTCTCAGCCAATCCCAGAATC
rs132709	ACGTTGGATGTGAGTCCTGTCCAAGATGAG	ACGTTGGATGAGCCCTTCCTAAGGATTCTG
rs132710	ACGTTGGATGTGAGTCCTGTCCAAGATGAG	ACGTTGGATGCCTAAGGATTCTGGGATTGG
rs132711	ACGTTGGATGTCATCTTGGACAGGACTCAG	ACGTTGGATGTTGCCATGGCAACCAAGTCA
rs132712	ACGTTGGATGGTCTTCAAGGCTGAGTGAGC	ACGTTGGATGACTCCACGTGGCCTCTCTTG
rs132713	ACGTTGGATGAAGGCTGAGTGAGCCCCAAC	ACGTTGGATGACTCCACGTGGCCTCTCTT
rs132714	ACGTTGGATGACACGGTGAAACCCCTTCTC	ACGTTGGATGAGTAGCTGGGACTACAGGTG
rs132716	ACGTTGGATGTGGATTTGCAATGAGGAGTC	ACGTTGGATGTCAATGACTGTGCTCTACTC
rs132717	ACGTTGGATGAATGTGGGCAGTTTTACGTG	ACGTTGGATGGATGGACCTTAGGGTGTTTC
rs132718	ACGTTGGATGTCAGAGGGTATCAACATCTC	ACGTTGGATGTGGGCATCTTCATATACTGC
rs132719	ACGTTGGATGATACCCTCAGTTGTACCCAG	ACGTTGGATGCTGAACAAAGGAGAAGGAGG
rs132720	ACGTTGGATGATACTGGGTACAACTGAGGG	ACGTTGGATGCCTCTCCACCTTTTCCTAAC
rs132721	ACGTTGGATGCAGCTACAAAGTTGCTAATGG	ACGTTGGATGTCTTATTTGTACCCTCCCTC
rs132722	ACGTTGGATGACAATGGTAATGCTTGGAGC	ACGTTGGATGGGGAGGGTACAAATAAGATG
rs132723	ACGTTGGATGACACAGATGTCTGTCTTCTG	ACGTTGGATGATACTCCCCTGGTGAATGCT
rs132724	ACGTTGGATGCCCCATGCAACAAGGGTAAA	ACGTTGGATGTGTCCCTTACAGCAAGAAGC
rs132725	ACGTTGGATGCCCCATGCAACAAGGGTAAA	ACGTTGGATGTGTCCCTTACAGCAAGAAGC

dbSNP rs#	Forward PCR primer	Reverse PCR primer
rs132726	ACGTTGGATGGGGTCACACAGTGAACAAAG	ACGTTGGATGTTCTTGCTGTAAGGGACAGG
rs132727	ACGTTGGATGGGGTCACACAGTGAACAAAG	ACGTTGGATGTTCTTGCTGTAAGGGACAGG
rs132728	ACGTTGGATGAGTTCTACTGGCTCATGGTG	ACGTTGGATGTTCGCCTTCTTCCTGCTTTG
rs132729	ACGTTGGATGAGTTCTACTGGCTCATGGTG	ACGTTGGATGTTCGCCTTCTTCCTGCTTTG
rs132730	ACGTTGGATGTTCGCCTTCTTCCTGCTTTG	ACGTTGGATGAGTTCTACTGGCTCATGGTG
rs132731	ACGTTGGATGTTTGTTCACTGTGTGACCCC	ACGTTGGATGTCCTGGTCCTCCCAGTTCTA
rs132732	ACGTTGGATGGCCAAGGCAACCATCTCAAC	ACGTTGGATGTGCAGCTCATCACAAGCGTC
rs132733	ACGTTGGATGGCAACCATCTCAACACCATG	ACGTTGGATGTGCAGCTCATCACAAGCGTC
rs140575	ACGTTGGATGTGCAGCTCATCACAAGCGTC	ACGTTGGATGCCATCTCAACACCATGAGCC
rs132734	ACGTTGGATGGGATACTGACTGTTAGCCTC	ACGTTGGATGCGGAATTGACCAACTGGTAG
rs132735	ACGTTGGATGGGATACTGACTGTTAGCCTC	ACGTTGGATGCGGAATTGACCAACTGGTAG
rs80587	ACGTTGGATGTCTGAGCCAAGCTCACCAGA	ACGTTGGATGTTTTCCTGCCCAAGGAGGAG
rs132736	ACGTTGGATGTTTTCCTGCCCAAGGAGGAG	ACGTTGGATGAAGCTCACCAGATGCAGACG
rs132737	ACGTTGGATGTCTAGGCAGCAATGAGCTAG	ACGTTGGATGTGCTCCTCCTGAGAAATCAC
rs132738	ACGTTGGATGTGAAGCCTGTAATCCCAGTG	ACGTTGGATGCATAGAGACAGCATCTCCTG
rs1807673	ACGTTGGATGTTGCAGTGAGCAGAGATTGC	ACGTTGGATGGTGAAATCTGAGTCGTGGTC
rs2014700	ACGTTGGATGTTGCAGTGAGCAGAGATTGC	ACGTTGGATGGTGAAATCTGAGTCGTGGTC
rs132739	ACGTTGGATGGGAATCAAAGAAGGTGGAGG	ACGTTGGATGTGGTTGTGGCCAGACCATAA
rs1812023	ACGTTGGATGAGCAGGAGGGAGGGAGCAAT	ACGTTGGATGGACCTCCCTCCATCTCCTTA
rs1812024	ACGTTGGATGTGTCAGAGGAAGATCCCTTG	ACGTTGGATGCCTCATAGAGCTATTGCGAG
rs2005590	ACGTTGGATGGGAGGCAATGCCTGATTTTG	ACGTTGGATGTGCTTCCACCACCTGGAAAA
rs132740	ACGTTGGATGTTCATTCTTCTTGTGCACAG	ACGTTGGATGCTTGACTGGTACCTAACAATG
rs3986001	ACGTTGGATGTTTTCTGACTTGGCATCACC	ACGTTGGATGCACAAAGTATTCCACCTTCC
rs2413390	ACGTTGGATGATAAATTCGTGGCTGAGCTC	ACGTTGGATGATCTTGTGGCATAAGGAGTC
rs132743	ACGTTGGATGACAGGGAGAAAGTGAGGAAG	ACGTTGGATGCATCCTGTTTCCCCTAAAGG
rs132744	ACGTTGGATGGGGTCTGTTTCAGGAGCATG	ACGTTGGATGTCTATGGCTGATGGCCACAG
rs2413391	ACGTTGGATGGGGCAAAAGCAGAAATACTG	ACGTTGGATGCCCTCAAACCCTGTTTTCTG
rs132749	ACGTTGGATGCCATGCACTCTCTAGTACTC	ACGTTGGATGTGTGGCCTTGGGGAAATGAT
rs132750	ACGTTGGATGTCCTGTGCCTGTGGAAACTC	ACGTTGGATGGGTTCTCCAGGTGGCAAAAG
rs132741	ACGTTGGATGCTACAATTTATCCGCACTAG	ACGTTGGATGGCCAAGTCAGAAAAATGAGAG
rs2413388	ACGTTGGATGTACAGAATTCAGACCAACCC	ACGTTGGATGGCCCTGAGATTTGATTTTCC
rs2413389	ACGTTGGATGGCTAGAATCTCATAACAGACG	ACGTTGGATGGCGTCCTACTATGATTTGTC

TABLE 12

dbSNP rs#	Extend Primer	Term Mix
rs3888818	TGAGACCAGCCTAGCCAAC	ACT
rs2010605	TGGTCCCAAGATATTCTATAGA	ACT
rs743919	CTCACCTAAGGACTGCCTCT	ACT
rs1008134	CCGGGCATCTTTTCTTCCATC	ACT
rs132607	GCAGGCAGGACAGCATGTG	ACT
rs1476029	CTTAGAGGCTATATTAAGACCA	ACG
rs1476030	TTTGGCAATACTGGCCTATTC	ACT
rs2413380	TCCAGGAGGGAAGACAACC	ACT
rs2051609	CGCTTGAGGTCAGGACCAG	CGT

dbSNP rs#	Extend Primer	Term Mix
rs2413381	GCGTGTTGCCAACAGCCTC	ACT
rs1894604	AAATGGGAGGGAATGTTGGC	ACG
rs1894605	TGCAGATCGCAACTGAGCG	ACT
rs132609	GTTTCTCAGAGGATCAGGGA	ACT
rs132610	GGTGTGAGATTTGGAGACTTT	ACG
rs132611	CTCTGTCCTCTAGCCCCC	ACT
rs132612	ATCTTCCCGCTACCTCAAGAGT	ACT
rs1008790	CATAATCACAAGTCCTATGATTA	ACT
rs23085	AGGATGACCATGGCAAGGAA	CGT
rs105161	GGCCCTGGCAGGAAACAG	ACG
rs132613	CCAATGGCCTCCACTGGC	ACT
rs132614	CGGCCACAGCGCTGCCC	ACG
rs132615	GCTTTCAGAACAACGGTAGAA	ACT
rs132617	AAGAGTGTGTGTGCAGTAGCAAG	ACG
rs3865724	TCACTTAAGCTTTGAATGTTTCTG	ACT
rs2019657	GCCAGAACATTGTGTTTCATTTGT	ACG
rs3865725	GGCAAGAGATACAGAATGCACA	CGT
rs2019364	GTATCTCTTGCCCCTGCTC	ACG
rs2008383	GAAGGACAGAAGGCTGATGC	ACG
rs3986002	TCCTTCTCTGTGTCACTCCT	CGT
rs3888942	ACATGGAAGCAGGGGTTTGA	CGT
rs3888943	AGCAGGGGTTTGATGAAATCT	ACT
rs3888944	CATCCTCCACATTGGGCCAA	ACT
rs132618	AATCTCAGCTGGAAGTGG	CGT
rs132619	TGCAACCAGCATTGACCG	CGT
rs3827346	GAGGCTGCACCATCTCCAA	ACG
rs132620	GGAAGCCTTTATTCAGGATTGT	ACT
rs132621	TCAAATCTGCAACTGGTGTCAGAA	ACT
rs80575	GAACTCTCAAGCCACTTGAC	ACT
rs80576	GCTAAGGCATTAGTTTGGCTGG	ACG
rs80577	TGAAATTGCACATGGCATTGG	ACG
rs80578	ATGCCTGGGAACTGGGGC	ACT
rs80579	GGGAGGCACTGAGGGCATGAAA	ACT
rs80580	CTGAGAATGAACAGCAGGTCA	ACT
rs132622	ATAGTAGTTCAATCAGATGGGC	ACT
rs132623	ATTCCAGCCTCTCTGTGTTCTG	ACT
rs132624	ACAGGCATGAGCCGCTGC	ACT
rs132625	GGCCACCGCATCCGGCTA	CGT
rs132626	AGTGGCATGATCTCGGCCCAC	ACG
rs132627	GAGGCGGAGGTTGCGGTG	ACT
rs1807672	CATTGAGAATAAGGTGGTTCGA	ACT
rs132628	CCTCCAAATTCATATACTGAGACC	ACG
rs132629	TCTCTCTCTCTCACACAC	CGT
rs132630	CTGCTCCTGGCTTACAGAG	ACT

dbSNP rs#	Extend Primer	Term Mix
rs132631	TCACAGTTCTGGAGGCAAAAA	ACT
rs132632	TGGAATTTTTGCCTCCAGAACTGT	ACT
rs132633	AGTGATTCACCAGGGAAGTGCCA	ACG
rs132634	GGTGCTTTGTGGAGGAACC	ACT
rs132635	ATTTCCCGTACATGGGGAGAAA	ACT
rs132636	TGACAATAGGCACATGGCAG	ACT
rs129603	GCTGCCATCCTAAACACATCTA	ACT
rs132637	ACACAGCAGGATTACTGCCCAGA	ACT
rs3788518	TGGGAGGCTCAAGGAAGAAACTCT	ACT
rs132638	GGAAAAAATAAAAGCAAAATACCC	ACT
rs132639	AAAGCAAACAGGCCTTCAGAA	CGT
rs132640	GGTGACACAGAGAAGACGTGGC	ACT
rs132641	GGGAGGTCAGAGGTCGGG	ACT
rs132642	GTCACTGAGAGACTTTCC	CGT
rs132643	GACACCCAGTACACACTGGCT	ACT
rs132644	TTTGGAATGAGGAGTCATTTACA	ACT
rs132645	CTCAACAGTAAGCAAGATTTAAA	ACT
rs2017329	TGATGTTCAGATTTTCCTTTTTTT	CGT
rs739198	TGGTCTCCACAACCTCTTATC	ACT
rs132647	AAACCATGGAAGTCTCTAGAGTCA	ACT
rs2097465	CAAACTGCAGGCTTGCCCAG	ACT
rs2105915	CAAGCCTGGGCAGCATAGCAC	ACT
rs132648	AATTCCCGTGCTAATGCACG	ACT
rs132649	CAGTCTTATTACTTTTGTACGAGG	ACT
rs132650	CATGAGCCACCGTGCCAG	ACT
rs132651	CCTCAGGGTTTTTCACCTGCCT	ACT
rs132652	AGGGCATCCTAACCCCCTA	CGT
rs80584	ATCTACCTGCTCAACTTCCTGA	ACG
rs132653	AATAACCAGACACGTTCTCCAG	ACT
rs916334	GTCCAGCAGCACCCTTGGT	ACG
rs132654	CGACAAGAGCAGGTCTGGAAC	ACT
rs132655	CAGAAGAACCCACATAAGGAA	ACT
rs132656	TCTTTGTCTTTTACTCCCACATCC	ACT
rs3834684	CACTAGAAGCAAGGAACCCCC	ACT
rs132657	TACCTGACAATCACCCCCC	CGT
rs916335	TCAGGTAATTGTCAGTCAGCC	ACT
rs132659	AGAACTCCCCAAATCGTCCT	ACG
rs132660	CCCCAGAGTGGGCTTTTCT	ACT
rs132661	CCGCTCTCCCTCTGAGAGT	ACT
rs132662	TGATCTGAGTTTACAGGTGAG	ACT
rs132663	TGAGATCTGTCTCAGACGCA	CGT
rs132664	CTGGGCTAGAGAGGGAGAC	ACT
rs132667	CTTTAACTTTTGCTCACAAGAGT	ACT
rs132670	AGACCAGCCTGATCAACATG	ACT

dbSNP rs#	Extend Primer	Term Mix
rs132671	ACATCAATAGGCCTAAAAATCGTT	ACT
rs132672	GAAACTTGAAATTCCTTGAGAAAT	ACT
rs132673	GTGTGAGTGGGAAGCCTCC	ACG
rs132674	GGAGTTGGAGGCTGTAGTAA	ACT
rs132675	AGATGTGCAATGGAATTTGGCAA	ACT
rs80585	AGGCATAATATTCTGACCATTAAG	ACG
rs80586	GGTCTGCTACTAGAATTCAGAA	ACG
rs132676	ATCCCTTAATATTGCATAGGAC	CGT
rs132677	GGGTTGAAGTACTATGCTAGT	CGT
rs132678	AGGTTAGTTCATGTAACTCCAT	ACT
rs132680	TTTTATTTTAGCTTGAGCTTTTCA	ACG
rs132681	CTAGCTCTAAATCACATTCTGC	ACG
rs132683	CAGGCCCATACCCAAAATATGCT	ACT
rs2269594	TGGCTAAAAGGACAGATAGAG	ACT
rs132684	GACACTAAGAGCGGTGAGAC	ACT
rs132685	CGTGCCACCCAACTGGAGA	ACT
rs132686	TGTGCATCTTATGGTGTACCA	ACG
rs132687	TCGTTACCCCCATTCTATCC	ACT
rs132688	ACAGGACGTGCTTGAAAGAG	ACG
rs132689	TGGCGATGGCCTCTGCTC	ACT
rs132690	GCTCTCCTTGCTTCAAAAAAAAA	CGT
rs132691	TGACCTATCCTGCTTCAGGT	ACT
rs132692	CGAAGTGTGTTAGCTCATGAC	ACT
rs132693	GAGTGTGGACACCAGGTCA	ACT
rs132694	CACCTTAGGAATGGCAGCTTC	ACG
rs132695	TATGACTACACATGCTGGCAAAC	ACT
rs1966266	GCAGACCCCTAACTCTAATTTG	ACG
rs1966267	CACTGAGTTATGAGTACTCAAC	ACT
rs106808	GATCCTTCCCGAGGACACC	ACT
rs132696	CAGGCTGCCTGGAAGGAGA	ACT
rs2239829	GATGGCTGGATTCATAACAGGTAA	ACT
rs2285154	AGTGCTGGGATTACAGGCAT	ACT
rs2239830	CAGTCACCTGAATTTGTGCTTATT	ACT
rs2239831	GTCACTGACCCAAGCTATCCTC	ACG
rs3865722	CAATTGCAAGCAACAGAAACAGAA	CGT
rs3865723	TCCACTGTACTGCTAGTATTG	ACG
rs3985996	AGAATTTCTCCCGGCATCAG	ACG
rs3985997	TCTCCCGGCATCAGCCTTC	ACG
rs3985998	GGGAGGAGTGACACAGAGAAGGA	ACG
rs3985999	GGGGTACTGGGAGGAGTGACAC	ACT
rs3986000	GCAGGACTGGGGTACTGG	ACT
rs2413382	CAGGGGAACTCAGGCCACA	ACT
rs2413383	AGCCATTGAAGACATGGAAGCC	ACT
rs2413384	ATTGAAGACATGGAAGCCGG	ACT

dbSNP rs#	Extend Primer	Term Mix
rs2413385	GGCCAGCTCTTCCTCCAC	CGT
rs2413386	CTTGGCCCAATGTGGAGGA	CGT
rs1894606	CCTAGCGGCAAGGGCTGT	ACT
rs916336	AGGGTATTTGCTGTCTGGCTGTCT	ACT
rs916337	AAGGGCCTGTATGTAGGTTGAA	ACT
rs916338	CCCTCTATGTCTCATGGATTTTCC	ACG
rs132697	CCTAGGGGAGCCCATATATCA	ACT
rs12781	AGACAGCTCGAGAGATCC	ACT
rs1053983	GCTGACTCAGATACACCCC	ACG
rs1053982	GATGCTGGTAAGACAGGG	ACG
rs2227167	CGTCAAAATCAAGTGCAAA	ACT
rs2227168	CATTAAGGGACATTCTGC	ACT
rs132700	ATCCTGTCTGTCATTGGCGTT	ACT
rs3075364	GTTTAGGGAGTGGTTTTTGAAAG	CGT
rs2227169	TGTCCTTTATTGGTACAGGGAAGA	ACT
rs2097466	TACAAGCATGAGCCACCGC	ACT
rs2097467	TCCCAAAGTGCTGGGATTACA	ACT
rs2413387	TACAGGCACTCACCACCAC	ACT
rs132701	TGTACAAAACATATTTAACCTTGA	ACT
rs132702	CACCCAGTTAAGGAAAAATTCCT	ACT
rs132703	GATCAATGTGTGTTCCCGGA	ACT
rs2269595	GCCCCAGACAGCATCTCC	ACT
rs2269596	TTGCTGGCAAGAGACCAGG	ACT
rs132704	GGGCTGCCTGGAGGAGG	ACG
rs2007468	TGGGCAATTCAGCCACACGCAC	ACT
rs132705	TATAGACTGAATTGTGTGCCC	ACG
rs2007706	GGAATTTACATAAGGGTCTATAG	ACT
rs132706	GAACCCCCTCCACTGCCC	ACT
rs132707	GCTCTCCCTCTGAAACAAGATG	ACT
rs132708	GAGAGCTTCTTCCTTGGCC	ACT
rs132709	CAGGGAAGATTAGAAGCTGAGAGC	CGT
rs132710	GGGCAGGGAAGATTAGAAGC	ACG
rs132711	TTTGCTGTCCAGGGCGGC	ACG
rs132712	AACCCAGACGGAGGTGGC	ACG
rs132713	GCCCCAACGGAACCCAGA	ACG
rs132714	CCCCTTCTCTACTGAAAATACAAA	ACT
rs132716	TGAGGAGTCATTTACCATGAG	ACG
rs132717	GCAGTTTTACGTGAAGGAGG	ACT
rs132718	GTTTTATACCTAGAGCCACACT	ACT
rs132719	TACCCAGTATTTCTTAACTTCC	CGT
rs132720	CAGCTACAAAGTTGCTAATGG	ACT
rs132721	ATGTTAGGAAAAGGTGGAGAG	ACT
rs132722	CCTAACTGGGATGGGCCTGAA	ACT
rs132723	TTCTGGGGCCCCCATGCA	ACT

dbSNP rs#	Extend Primer	Term Mix
rs132724	ACCCAGTCCTGGGCAGCA	ACT
rs132725	GGGAGTATGCAGAGGGGC	ACT
rs132726	CAGTGAACAAAGCAGGAAGAAGG	ACT
rs132727	GTCACACAGTGAACAAAGCAGGAA	ACT
rs132728	CATGGTGTTGAGATGGTTGCC	ACG
rs132729	GCTCATGGTGTTGAGATGGTT	ACT
rs132730	GTGACCCCCTAGGCCAAGGCA	ACT
rs132731	GGCAACCATCTCAACACCAT	ACT
rs132732	AACCATCTCAACACCATGAGCCA	ACT
rs132733	ATCTCAACACCATGAGCCAG	ACT
rs140575	GCTCTCTCCTGGTCCTCC	ACG
rs132734	AGCCTCAACTAGGACACA	ACT
rs132735	TCAACTAGGACACAGTGC	ACT
rs80587	GCCAAGCTCACCAGATGCAGA	ACT
rs132736	AGGGAGCTGCTTTGCTGAAA	ACT
rs132737	CCTGCAGCCTGGGTGACA	ACT
rs132738	GGCCAGGAGTTCAAGACAGCCTG	ACT
rs1807673	GGGCAACAGAGCGAGACTCC	ACT
rs2014700	GAGCGAGACTCCATCTCA	ACT
rs132739	GGGAGGTGACCTGGAGCC	ACG
rs1812023	GCAATCAGACTCAAGCCTGG	ACT
rs1812024	GGGATGGTGTGACCTCCC	ACG
rs2005590	CAATGCCTGATTTTGTCACTGAAC	ACT
rs132740	GGCATATGTGCATTTGTCTGAG	ACG
rs3986001	TCCTTTTTCTAAACCCCTGCAA	ACT
rs2413390	GGCTGAGCTCAAGGTTTTAAA	ACT
rs132743	GGAGAAAGTGAGGAAGAAAATTA	ACT
rs132744	TGGGGTTACAGTTGGTCATAACC	ACT
rs2413391	TGATATGTTCAGCGGTGCAC	ACT
rs132749	TCTTGATGTTTCTCCTATCCC	ACG
rs132750	CCTGTGGAAACTCAGCAGC	ACG
rs132741	GCACTAGATATTGAATTCTTTCC	ACT
rs2413388	CAACCCCGTGACTGGAGATTC	CGT
rs2413389	TTTCTCTCTCTAGTACTCTATTT	ACT

Genetic Analysis

[0229] Allelotyping results from the discovery cohort are shown for cases and controls in Table 13. The allele frequency for the A2 allele is noted in the fifth and sixth columns for osteoarthritis case pools and control pools, respectively, where "AF" is allele frequency. The allele frequency for the A1 allele can be easily calculated by subtracting the A2 allele frequency from 1 (A1 AF = 1-A2 AF). For example, the SNP rs2010605 has the following case and control allele frequencies: case A1 (A) = 0.19; case A2

(G) = 0.81; control A1 (A) = 0.18; and control A2 (G) = 0.82, where the nucleotide is provided in paranthesis. Some SNPs are labeled "untyped" because of failed assays.

TABLE 13

dbSNP rs#	Position in SEQ ID NO: 1	Chromosome Position	A1/A2 Allele	F A2 Case AF	F A2 Control AF	F p-Value
rs3888818	201	34781551	C/T			
rs2010605	425	34781775	A/G	0.81	0.82	0.782
rs743919	1095	34782445	G/T	0.10	0.11	0.502
rs1008134	2201	34783551	A/C			
rs132607	7879	34789229	A/G	0.11	0.11	0.813
rs1476029	8395	34789745	C/T	0.15	0.15	0.983
rs1476030	8461	34789811	C/T	0.36	0.37	0.708
rs2413380	9503	34790853	C/T	0.29	0.29	0.900
rs2051609	10304	34791654	G/T			
rs2413381	10695	34792045	C/T			
rs1894604	16300	34797650	A/G	0.08	0.08	0.759
rs1894605	16444	34797794	G/T	0.08	0.09	0.468
rs132609	17591	34798941	C/T	0.68	0.67	0.777
rs132610	17988	34799338	-/A			
rs132611	19116	34800466	-/T	0.14	0.15	0.863
rs132612	19358	34800708	C/T	0.23	0.23	0.951
rs1008790	20300	34801650	A/G	0.03	0.10	**0.007**
rs23085	20669	34802019	A/T	0.31	0.32	0.738
rs105161	20891	34802241	A/G	0.76	0.77	0.826
rs132613	21451	34802801	C/T	0.80	0.81	0.619
rs132614	21978	34803328	C/T	0.16	0.14	0.434
rs132615	22785	34804135	C/G	0.32	0.31	0.740
rs132617	24248	34805598	C/T	0.35	0.36	0.825
rs3865724	24770	34806120	C/T	0.65	0.65	0.940
rs2019657	24844	34806194	A/G	0.20	0.20	0.857
rs3865725	25066	34806416	G/T			
rs2019364	25096	34806446	C/T	0.40	0.39	0.767
rs2008383	25309	34806659	A/G	0.18	0.17	0.665
rs3986002	25344	34806694	A/C			
rs3888942	25529	34806879	A/T			
rs3888943	25537	34806887	A/G			
rs3888944	25554	34806904	A/C			
rs132618	27963	34809313	A/T	0.43	0.43	0.934
rs132619	28134	34809484	G/T			
rs3827346	28356	34809706	A/G	0.84	0.84	0.976
rs132620	29648	34810998	-/A	0.29	0.29	0.879
rs132621	29986	34811336	A/G	0.32	0.31	0.867
rs80575	30217	34811567	G/T	0.27	0.27	0.948
rs80576	30267	34811617	A/G	0.26	0.25	0.443
rs80577	30315	34811665	A/G	0.26	0.23	0.191
rs80578	30585	34811935	C/T	0.49	0.48	0.548
rs80579	30724	34812074	A/C	0.23	0.25	0.574
rs80580	30897	34812247	C/T	0.31	0.31	0.878
rs132622	30931	34812281	C/T	0.29	0.30	0.943
rs132623	31080	34812430	G/T	0.60	0.59	0.806
rs132624	31246	34812596	C/T	0.36	0.37	0.772
rs132625	31373	34812723	A/T			
rs132626	31463	34812813	A/G	0.89	0.84	0.082
rs132627	31467	34812817	A/G	0.12	0.11	0.836
rs1807672	32188	34813538	G/T	0.30	0.30	0.974
rs132628	32288	34813638	C/T	0.25	0.24	0.691
rs132629	32520	34813870	A/T	0.04	0.06	0.250

dbSNP rs#	Position in SEQ ID NO: 1	Chromosome Position	A1/A2 Allele	F A2 Case AF	F A2 Control AF	F p-Value
rs132630	32594	34813944	A/C	0.75	0.75	0.978
rs132631	32657	34814007	A/C	0.72	0.73	0.509
rs132632	32677	34814027	A/G	0.66	0.65	0.798
rs132633	32764	34814114	C/T	0.34	0.33	0.796
rs132634	32784	34814134	A/G	0.45	0.42	0.317
rs132635	32830	34814180	C/T	0.41	0.40	0.772
rs132636	32872	34814222	C/T	0.41	0.44	0.192
rs129603	33121	34814471	A/C			
rs132637	33348	34814698	G/T	0.09	0.09	0.628
rs3788518	33952	34815302	C/G	0.17	0.19	0.297
rs132638	34184	34815534	C/G	0.56	0.58	0.509
rs132639	34361	34815711	A/T	0.13	0.12	0.561
rs132640	35026	34816376	A/G	0.32	0.30	0.388
rs132641	35192	34816542	A/G	0.48	0.51	0.287
rs132642	35600	34816950	A/T	0.15	0.14	0.732
rs132643	36033	34817383	C/T	0.44	0.46	0.360
rs132644	36289	34817639	C/T	0.53	0.58	0.075
rs132645	38869	34820219	A/G	0.19	0.18	0.572
rs2017329	39629	34820979	A/T	0.39	0.40	0.915
rs739198	40530	34821880	C/T	0.70	0.70	0.878
rs132647	41621	34822971	C/T	0.23	0.23	0.957
rs2097465	42379	34823729	C/T	0.54	0.51	0.344
rs2105915	42802	34824152	C/T	0.57	0.56	0.468
rs132648	42865	34824215	T/C			
rs132649	43644	34824994	A/G	0.21	0.22	0.579
rs132650	45051	34826401	C/T	0.34	0.31	0.248
rs132651	45828	34827178	A/C			
rs132652	45829	34827179	A/T			
rs80584	46257	34827607	C/T	0.81	0.75	**0.043**
rs132653	47286	34828636	A/C	0.17	0.15	0.312
rs916334	47427	34828777	C/T	0.34	0.36	0.345
rs132654	47963	34829313	C/T	0.54	0.56	0.439
rs132655	48013	34829363	C/T	0.41	0.41	0.838
rs132656	48229	34829579	C/T	0.37	0.36	0.813
rs3834684	48282	34829632	-/A	0.21	0.22	0.480
rs132657	48376	34829726	-/G	0.49	0.50	0.719
rs916335	48404	34829754	A/G	0.38	0.36	0.509
rs132659	49900	34831250	C/T			
rs132660	52699	34834049	G/T	0.37	0.38	0.754
rs132661	52897	34834247	A/G	0.27	0.29	0.291
rs132662	53414	34834764	A/G	0.61	0.58	0.186
rs132663	53487	34834837	A/T	0.26	0.29	0.167
rs132664	54112	34835462	G/T	0.25	0.29	0.098
rs132667	55492	34836842	A/G	0.35	0.36	0.559
rs132670	59766	34841116	C/T			
rs132671	60307	34841657	A/G	0.49	0.53	0.145
rs132672	60701	34842051	A/G	0.23	0.22	0.716
rs132673	60952	34842302	A/G	0.41	0.37	0.184
rs132674	61401	34842751	C/T	0.32	0.31	0.476
rs132675	62379	34843729	C/T	0.38	0.35	0.188
rs80585	62870	34844220	C/T	0.28	0.27	0.525
rs80586	62879	34844229	A/G	0.66	0.66	0.966
rs132676	63499	34844849	A/T	0.26	0.23	0.177
rs132677	64284	34845634	-/A	0.69	0.69	0.873
rs132678	64408	34845758	A/G	0.46	0.48	0.395
rs132680	64760	34846110	A/G	0.20	0.20	0.995
rs132681	65230	34846580	A/G	0.24	0.24	0.901

dbSNP rs#	Position in SEQ ID NO: 1	Chromosome Position	A1/A2 Allele	F A2 Case AF	F A2 Control AF	F p-Value
rs132683	66127	34847477	A/G	0.19	0.19	0.851
rs2269594	66634	34847984	C/T	0.70	0.67	0.332
rs132684	66686	34848036	A/G	0.21	0.20	0.756
rs132685	66694	34848044	C/G	0.30	0.28	0.553
rs132686	67113	34848463	A/G	0.46	0.48	0.398
rs132687	67257	34848607	A/G	0.96	0.96	0.767
rs132688	67403	34848753	A/G	0.24	0.23	0.553
rs132689	67609	34848959	A/G	0.61	0.63	0.564
rs132690	68418	34849768	-/A	0.16	0.17	0.672
rs132691	68610	34849960	C/G	0.52	0.52	0.976
rs132692	69629	34850979	C/T	0.63	0.62	0.800
rs132693	70024	34851374	A/G	0.58	0.58	0.868
rs132694	70848	34852198	A/G	0.17	0.16	0.583
rs132695	71428	34852778	C/G	0.23	0.22	0.616
rs1966266	71553	34852903	C/T	0.49	0.47	0.413
rs1966267	71633	34852983	A/G	0.40	0.41	0.773
rs106808	71768	34853118	A/C	0.68	0.67	0.617
rs132696	71769	34853119	A/G			
rs2239829	73039	34854389	A/G	0.34	0.36	0.510
rs2285154	73325	34854675	A/G			
rs2239830	73412	34854762	A/C	0.49	0.50	0.841
rs2239831	73547	34854897	C/T	0.52	0.50	0.564
rs3865722	73769	34855119	A/T	0.57	0.56	0.861
rs3865723	73806	34855156	A/G	0.59	0.58	0.722
rs3985996	74467	34855817	C/T	0.30	0.29	0.861
rs3985997	74472	34855822	C/T	0.89	0.90	0.527
rs3985998	74473	34855823	A/G			
rs3985999	74482	34855832	C/T	0.19	0.19	0.968
rs3986000	74494	34855844	A/C			
rs2413382	74592	34855942	A/G	0.61	0.59	0.618
rs2413383	74670	34856020	G/T			
rs2413384	74672	34856022	G/T			
rs2413385	74714	34856064	G/T	0.70	0.70	0.944
rs2413386	74723	34856073	A/T	0.70	0.71	0.816
rs1894606	74749	34856099	A/G			
rs916336	74861	34856211	C/G			
rs916337	74892	34856242	C/T			
rs916338	74893	34856243	C/T	0.40	0.40	0.939
rs132697	75176	34856526	A/G	0.59	0.61	0.418
rs12781	75705	34857055	A/G			
rs1053983	75989	34857339	A/G	0.42	0.43	0.848
rs1053982	76027	34857377	A/G	0.69	0.69	0.981
rs2227167	77949	34859299	A/G	0.35	0.37	0.392
rs2227168	77974	34859324	C/T	0.38	0.38	0.797
rs132700	78167	34859517	C/T	0.25	0.25	0.729
rs3075364	78310	34859660	-/CT	0.42	0.42	0.879
rs2227169	78415	34859765	C/T	0.38	0.37	0.619
rs2097466	78575	34859925	C/T	0.52	0.52	0.962
rs2097467	78590	34859940	C/T			
rs2413387	78709	34860059	C/T			
rs132701	78875	34860225	C/T	0.23	0.23	0.951
rs132702	79864	34861214	C/T			
rs132703	81316	34862666	C/T	0.52	0.52	0.995
rs2269595	81320	34862670	A/G	0.19	0.18	0.464
rs2269596	81409	34862759	C/T	0.40	0.40	0.753
rs132704	81737	34863087	C/T	0.65	0.66	0.906
rs2007468	81843	34863193	A/G	0.40	0.40	0.987

dbSNP rs#	Position in SEQ ID NO: 1	Chromosome Position	A1/A2 Allele	F A2 Case AF	F A2 Control AF	F p-Value
rs132705	82102	34863452	C/T	0.30	0.30	0.944
rs2007706	82833	34864183	C/T	0.32	0.32	0.944
rs132706	83461	34864811	C/T			
rs132707	83624	34864974	C/T	0.26	0.28	0.540
rs132708	83660	34865010	C/G	0.30	0.29	0.506
rs132709	83701	34865051	A/T	0.45	0.46	0.538
rs132710	83708	34865058	A/G	0.62	0.60	0.310
rs132711	83782	34865132	C/T			
rs132712	85707	34867057	A/G	0.84	0.84	0.693
rs132713	85717	34867067	A/G	0.29	0.30	0.911
rs132714	86486	34867836	C/T			
rs132716	86833	34868183	A/G			
rs132717	87115	34868465	C/T	0.48	0.49	0.647
rs132718	87234	34868584	A/G	0.68	0.69	0.453
rs132719	87479	34868829	G/T	0.59	0.54	0.078
rs132720	87561	34868911	A/G	0.12	0.13	0.607
rs132721	87604	34868954	A/G	0.73	0.73	0.919
rs132722	87674	34869024	C/T	0.22	0.22	0.985
rs132723	87958	34869308	A/G	0.03	NA	**0.033**
rs132724	87992	34869342	-/G	0.12	0.12	0.830
rs132725	88019	34869369	A/G	0.69	0.64	0.097
rs132726	88074	34869424	C/G	0.10	0.13	0.240
rs132727	88079	34869429	C/G	0.68	0.67	0.865
rs132728	88115	34869465	A/G			
rs132729	88118	34869468	C/G			
rs132730	88120	34869470	A/G			
rs132731	88135	34869485	-/CTCAT			
rs132732	88142	34869492	G/T			
rs132733	88143	34869493	G/T			
rs140575	88149	34869499	ACA/TG	0.37	0.38	0.656
rs132734	88340	34869690	A/G	0.46	0.44	0.522
rs132735	88344	34869694	G/T	0.85	0.85	0.781
rs80587	88512	34869862	C/G	0.40	0.42	0.493
rs132736	88521	34869871	C/T			
rs132737	88650	34870000	C/G	0.32	0.33	0.824
rs132738	88827	34870177	C/T	0.78	0.80	0.327
rs1807673	89230	34870580	A/G	0.28	0.28	0.969
rs2014700	89236	34870586	A/G	0.89	0.92	**0.036**
rs132739	90754	34872104	G/A			
rs1812023	90984	34872334	A/G	0.70	0.70	0.997
rs1812024	91110	34872460	A/G	0.68	0.69	0.755
rs2005590	92026	34873376	C/T	0.71	0.73	0.252
rs132740	92954	34874304	C/T	0.09	0.08	0.370
rs3986001	93375	34874725	-/TTGC	0.51	0.54	0.302
rs2413390	93794	34875144	C/T	0.37	0.41	0.103
rs132743	94937	34876287	C/G	0.07	0.07	0.985
rs132744	95068	34876418	C/T	0.73	0.78	0.078
rs2413391	96188	34877538	A/G	0.39	0.40	0.938
rs132749	97092	34878442	C/T	0.47	0.50	0.329
rs132750	98812	34880162	C/T	0.67	0.67	0.710
rs132741	not mapped	not mapped	A/C	0.29	0.30	0.608
rs2413388	not mapped	not mapped	A/T	0.31	0.31	0.967
rs2413389	not mapped	not mapped	C/G	0.30	0.28	0.393

[0230] The *APOL3* proximal SNPs were also allelotyped in the replication cohorts using the methods described herein and the primers provided in Tables 11 and 12. The replication allelotyping results for replication cohort #1 and replication cohort #2 are provided in Tables 14 and 15, respectively.

TABLE 14

dbSNP rs#	Position in SEQ ID NO: 1	Chromosome Position	A1/A2 Allele	F A2 Case AF	F A2 Control AF	F p-Value
rs3888818	201	34781551	C/T			
rs2010605	425	34781775	A/G	0.79	0.80	0.664
rs743919	1095	34782445	G/T	0.11	0.11	0.982
rs1008134	2201	34783551	A/C			
rs132607	7879	34789229	A/G	0.10	0.11	0.589
rs1476029	8395	34789745	C/T	0.16	0.13	0.177
rs1476030	8461	34789811	C/T	0.33	0.38	0.149
rs2413380	9503	34790853	C/T	0.26	0.30	0.174
rs2051609	10304	34791654	G/T			
rs2413381	10695	34792045	C/T			
rs1894604	16300	34797650	A/G	0.07	0.08	0.476
rs1894605	16444	34797794	G/T	0.07	0.09	0.269
rs132609	17591	34798941	C/T	0.68	0.67	0.687
rs132610	17988	34799338	-/A			
rs132611	19116	34800466	-/T	0.14	0.14	0.992
rs132612	19358	34800708	C/T	0.22	0.22	0.799
rs1008790	20300	34801650	A/G	0.03	0.06	0.358
rs23085	20669	34802019	A/T	0.31	0.30	0.868
rs105161	20891	34802241	A/G	0.75	0.80	0.058
rs132613	21451	34802801	C/T	0.82	0.81	0.725
rs132614	21978	34803328	C/T	0.15	0.15	0.914
rs132615	22785	34804135	C/G	0.32	0.32	0.885
rs132617	24248	34805598	C/T	0.36	0.37	0.706
rs3865724	24770	34806120	C/T	0.66	0.64	0.473
rs2019657	24844	34806194	A/G	0.21	0.21	0.898
rs3865725	25066	34806416	G/T			
rs2019364	25096	34806446	C/T	0.41	0.39	0.480
rs2008383	25309	34806659	A/G	0.17	0.16	0.721
rs3986002	25344	34806694	A/C			
rs3888942	25529	34806879	A/T			
rs3888943	25537	34806887	A/G			
rs3888944	25554	34806904	A/C			
rs132618	27963	34809313	A/T	0.42	0.45	0.522
rs132619	28134	34809484	G/T			
rs3827346	28356	34809706	A/G	0.85	0.84	0.703
rs132620	29648	34810998	-/A	0.29	0.28	0.917
rs132621	29986	34811336	A/G	0.32	0.31	0.864
rs80575	30217	34811567	G/T	0.27	0.27	0.909
rs80576	30267	34811617	A/G	0.30	0.22	**0.007**
rs80577	30315	34811665	A/G	0.24	0.23	0.853
rs80578	30585	34811935	C/T	0.49	0.49	0.884
rs80579	30724	34812074	A/C	0.24	0.25	0.761
rs80580	30897	34812247	C/T	0.30	0.30	0.815
rs132622	30931	34812281	C/T	0.29	0.28	0.884
rs132623	31080	34812430	G/T	0.62	0.63	0.633
rs132624	31246	34812596	C/T	0.36	0.35	0.769
rs132625	31373	34812723	A/T			
rs132626	31463	34812813	A/G	0.92	NA	NA
rs132627	31467	34812817	A/G	0.10	0.10	0.844
rs1807672	32188	34813538	G/T	0.29	0.29	0.954

dbSNP rs#	Position in SEQ ID NO: 1	Chromosome Position	A1/A2 Allele	F A2 Case AF	F A2 Control AF	F p-Value
rs132628	32288	34813638	C/T	0.24	0.25	0.796
rs132629	32520	34813870	A/T	0.03	0.06	0.345
rs132630	32594	34813944	A/C	0.75	0.72	0.388
rs132631	32657	34814007	A/C	0.70	0.77	**0.017**
rs132632	32677	34814027	A/G	0.67	0.63	0.275
rs132633	32764	34814114	C/T	0.34	0.32	0.558
rs132634	32784	34814134	A/G	0.44	0.41	0.435
rs132635	32830	34814180	C/T	0.39	NA	0.409
rs132636	32872	34814222	C/T	0.42	0.46	0.179
rs129603	33121	34814471	A/C			
rs132637	33348	34814698	G/T	0.08	0.10	0.462
rs3788518	33952	34815302	C/G	0.19	0.19	0.855
rs132638	34184	34815534	C/G	0.54	0.60	0.114
rs132639	34361	34815711	A/T	0.13	0.09	0.153
rs132640	35026	34816376	A/G	0.31	0.32	0.767
rs132641	35192	34816542	A/G	0.46	0.53	0.074
rs132642	35600	34816950	A/T	0.16	0.12	0.101
rs132643	36033	34817383	C/T	0.42	0.48	0.081
rs132644	36289	34817639	C/T	0.50	0.62	**~0.0001**
rs132645	38869	34820219	A/G	0.21	0.15	**0.027**
rs2017329	39629	34820979	A/T	0.40	0.38	0.737
rs739198	40530	34821880	C/T	0.74	NA	0.697
rs132647	41621	34822971	C/T	0.21	0.24	0.430
rs2097465	42379	34823729	C/T	0.54	0.52	0.670
rs2105915	42802	34824152	C/T	0.57	0.55	0.582
rs132648	42865	34824215	T/C			
rs132649	43644	34824994	A/G	0.20	0.23	0.213
rs132650	45051	34826401	C/T	0.36	0.29	**0.033**
rs132651	45828	34827178	A/C			
rs132652	45829	34827179	A/T			
rs80584	46257	34827607	C/T	0.81	0.75	**0.043**
rs132653	47286	34828636	A/C	0.18	0.13	0.060
rs916334	47427	34828777	C/T	0.32	0.38	0.064
rs132654	47963	34829313	C/T	0.55	0.54	0.769
rs132655	48013	34829363	C/T	0.42	0.39	0.390
rs132656	48229	34829579	C/T	0.38	0.35	0.429
rs3834684	48282	34829632	-/A	0.21	0.23	0.416
rs132657	48376	34829726	-/G	0.47	0.53	0.089
rs916335	48404	34829754	A/G	0.39	0.33	0.153
rs132659	49900	34831250	C/T			
rs132660	52699	34834049	G/T	0.39	0.36	0.227
rs132661	52897	34834247	A/G	0.25	0.31	0.086
rs132662	53414	34834764	A/G	0.63	0.57	0.083
rs132663	53487	34834837	A/T	0.25	0.31	0.051
rs132664	54112	34835462	G/T	0.23	0.29	**0.037**
rs132667	55492	34836842	A/G	0.34	0.32	0.546
rs132670	59766	34841116	C/T			
rs132671	60307	34841657	A/G	0.49	0.55	**0.041**
rs132672	60701	34842051	A/G	0.24	0.20	0.211
rs132673	60952	34842302	A/G	0.40	0.36	0.284
rs132674	61401	34842751	C/T	0.32	0.29	0.453
rs132675	62379	34843729	C/T	0.38	0.33	0.178
rs80585	62870	34844220	C/T	0.29	0.24	0.164
rs80586	62879	34844229	A/G	0.65	0.69	0.212
rs132676	63499	34844849	A/T	0.26	0.20	**0.034**
rs132677	64284	34845634	-/A	0.69	0.72	0.438
rs132678	64408	34845758	A/G	0.48	0.50	0.572

dbSNP rs#	Position in SEQ ID NO: 1	Chromosome Position	A1/A2 Allele	F A2 Case AF	F A2 Control AF	F p-Value
rs132680	64760	34846110	A/G	0.20	0.18	0.429
rs132681	65230	34846580	A/G	0.25	0.22	0.362
rs132683	66127	34847477	A/G	0.19	0.17	0.449
rs2269594	66634	34847984	C/T	0.70	0.65	0.234
rs132684	66686	34848036	A/G	0.21	0.18	0.349
rs132685	66694	34848044	C/G	0.30	0.28	0.528
rs132686	67113	34848463	A/G	0.46	0.50	0.207
rs132687	67257	34848607	A/G	0.95	0.98	0.217
rs132688	67403	34848753	A/G	0.26	0.23	0.367
rs132689	67609	34848959	A/G	0.60	0.63	0.353
rs132690	68418	34849768	-/A	0.16	0.15	0.901
rs132691	68610	34849960	C/G	0.51	0.53	0.573
rs132692	69629	34850979	C/T	0.62	0.64	0.566
rs132693	70024	34851374	A/G	0.58	0.62	0.227
rs132694	70848	34852198	A/G	0.17	0.14	0.195
rs132695	71428	34852778	C/G	0.24	0.22	0.560
rs1966266	71553	34852903	C/T	0.49	0.46	0.389
rs1966267	71633	34852983	A/G	0.40	0.43	0.304
rs106808	71768	34853118	A/C	0.66	0.70	0.243
rs132696	71769	34853119	A/G			
rs2239829	73039	34854389	A/G	0.34	0.38	0.264
rs2285154	73325	34854675	A/G			
rs2239830	73412	34854762	A/C	0.49	NA	0.490
rs2239831	73547	34854897	C/T	0.52	0.48	0.254
rs3865722	73769	34855119	A/T	0.57	0.56	0.739
rs3865723	73806	34855156	A/G	0.58	0.57	0.819
rs3985996	74467	34855817	C/T	0.30	0.31	0.799
rs3985997	74472	34855822	C/T	0.88	0.89	0.941
rs3985998	74473	34855823	A/G			
rs3985999	74482	34855832	C/T	0.20	0.20	0.871
rs3986000	74494	34855844	A/C			
rs2413382	74592	34855942	A/G	0.61	0.58	0.311
rs2413383	74670	34856020	G/T			
rs2413384	74672	34856022	G/T			
rs2413385	74714	34856064	G/T	0.70	0.68	0.509
rs2413386	74723	34856073	A/T	0.71	0.70	0.807
rs1894606	74749	34856099	A/G			
rs916336	74861	34856211	C/G			
rs916337	74892	34856242	C/T			
rs916338	74893	34856243	C/T	0.41	0.39	0.465
rs132697	75176	34856526	A/G	0.58	0.63	0.101
rs12781	75705	34857055	A/G			
rs1053983	75989	34857339	A/G	0.41	0.45	0.301
rs1053982	76027	34857377	A/G	0.66	0.66	0.959
rs2227167	77949	34859299	A/G	0.35	0.39	0.313
rs2227168	77974	34859324	C/T	0.39	0.35	0.234
rs132700	78167	34859517	C/T	0.26	0.24	0.652
rs3075364	78310	34859660	-/CT	0.42	0.45	0.287
rs2227169	78415	34859765	C/T	0.40	0.36	0.293
rs2097466	78575	34859925	C/T	0.52	0.49	0.376
rs2097467	78590	34859940	C/T			
rs2413387	78709	34860059	C/T			
rs132701	78875	34860225	C/T	0.23	0.21	0.499
rs132702	79864	34861214	C/T			
rs132703	81316	34862666	C/T	0.52	0.52	0.995
rs2269595	81320	34862670	A/G	0.20	0.17	0.210
rs2269596	81409	34862759	C/T	0.42	0.37	0.172

dbSNP rs#	Position in SEQ ID NO: 1	Chromosome Position	A1/A2 Allele	F A2 Case AF	F A2 Control AF	F p-Value
rs132704	81737	34863087	C/T	0.64	0.67	0.398
rs2007468	81843	34863193	A/G	0.39	0.42	0.354
rs132705	82102	34863452	C/T	0.29	0.28	0.834
rs2007706	82833	34864183	C/T	0.29	0.27	0.546
rs132706	83461	34864811	C/T			
rs132707	83624	34864974	C/T	0.26	0.29	0.252
rs132708	83660	34865010	C/G	0.29	0.28	0.584
rs132709	83701	34865051	A/T	0.46	0.47	0.748
rs132710	83708	34865058	A/G	0.62	0.60	0.477
rs132711	83782	34865132	C/T			
rs132712	85707	34867057	A/G	0.82	0.82	0.984
rs132713	85717	34867067	A/G	0.31	0.33	0.416
rs132714	86486	34867836	C/T			
rs132716	86833	34868183	A/G			
rs132717	87115	34868465	C/T	0.48	NA	NA
rs132718	87234	34868584	A/G	0.67	0.71	0.237
rs132719	87479	34868829	G/T	0.56	0.52	0.254
rs132720	87561	34868911	A/G	0.12	0.14	0.416
rs132721	87604	34868954	A/G	0.72	0.71	0.626
rs132722	87674	34869024	C/T	0.23	0.24	0.791
rs132723	87958	34869308	A/G	NA	NA	
rs132724	87992	34869342	-/G	0.12	0.12	0.830
rs132725	88019	34869369	A/G	0.60	NA	0.691
rs132726	88074	34869424	C/G	0.09	0.15	0.103
rs132727	88079	34869429	C/G	0.68	0.66	0.536
rs132728	88115	34869465	A/G			
rs132729	88118	34869468	C/G			
rs132730	88120	34869470	A/G			
rs132731	88135	34869485	-/CTCAT			
rs132732	88142	34869492	G/T			
rs132733	88143	34869493	G/T			
rs140575	88149	34869499	ACA/TG	0.43	0.44	0.792
rs132734	88340	34869690	A/G	0.45	0.41	0.263
rs132735	88344	34869694	G/T	0.83	0.82	0.629
rs80587	88512	34869862	C/G	0.42	0.47	0.152
rs132736	88521	34869871	C/T			
rs132737	88650	34870000	C/G	0.33	0.35	0.631
rs132738	88827	34870177	C/T	0.74	0.79	0.131
rs1807673	89230	34870580	A/G	0.30	0.32	0.583
rs2014700	89236	34870586	A/G	0.85	0.90	0.043
rs132739	90754	34872104	G/A			
rs1812023	90984	34872334	A/G	0.70	NA	0.704
rs1812024	91110	34872460	A/G	0.66	0.68	0.563
rs2005590	92026	34873376	C/T	0.72	0.75	0.268
rs132740	92954	34874304	C/T	0.10	0.06	0.085
rs3986001	93375	34874725	-/TTGC	0.49	0.53	0.341
rs2413390	93794	34875144	C/T	0.35	0.40	0.143
rs132743	94937	34876287	C/G	0.06	0.05	0.608
rs132744	95068	34876418	C/T	0.72	0.78	0.069
rs2413391	96188	34877538	A/G	0.37	0.39	0.544
rs132749	97092	34878442	C/T	0.48	0.51	0.311
rs132750	98812	34880162	C/T	0.65	0.67	0.678
rs132741	not mapped	not mapped	A/C	0.28	0.30	0.509
rs2413388	not mapped	not mapped	A/T	0.32	0.32	0.908
rs2413389	not mapped	not mapped	C/G	0.32	0.29	0.480

TABLE 15

dbSNP rs#	Position in SEQ ID NO: 1	Chromosome Position	A1/A2 Allele	F A2 Case AF	F A2 Control AF	F p-Value
rs3888818	201	34781551	C/T			
rs2010605	425	34781775	A/G	0.84	0.84	0.924
rs743919	1095	34782445	G/T	0.08	0.11	0.294
rs1008134	2201	34783551	A/C			
rs132607	7879	34789229	A/G	0.13	0.10	0.297
rs1476029	8395	34789745	C/T	0.13	0.18	0.126
rs1476030	8461	34789811	C/T	0.40	0.35	0.214
rs2413380	9503	34790853	C/T	0.33	0.28	0.163
rs2051609	10304	34791654	G/T			
rs2413381	10695	34792045	C/T			
rs1894604	16300	34797650	A/G	0.09	0.08	0.726
rs1894605	16444	34797794	G/T	0.09	0.09	0.834
rs132609	17591	34798941	C/T	0.67	0.67	0.950
rs132610	17988	34799338	-/A			
rs132611	19116	34800466	-/T	0.14	0.15	0.749
rs132612	19358	34800708	C/T	0.23	0.24	0.780
rs1008790	20300	34801650	A/G	untyped	0.16	NA
rs23085	20669	34802019	A/T	0.32	0.35	0.395
rs105161	20891	34802241	A/G	0.78	0.71	**0.037**
rs132613	21451	34802801	C/T	0.78	0.82	0.245
rs132614	21978	34803328	C/T	0.17	0.13	0.167
rs132615	22785	34804135	C/G	0.31	0.30	0.691
rs132617	24248	34805598	C/T	0.35	0.35	0.862
rs3865724	24770	34806120	C/T	0.64	0.67	0.384
rs2019657	24844	34806194	A/G	0.17	0.18	0.745
rs3865725	25066	34806416	G/T			
rs2019364	25096	34806446	C/T	0.38	0.39	0.702
rs2008383	25309	34806659	A/G	0.18	0.18	0.853
rs3986002	25344	34806694	A/C			
rs3888942	25529	34806879	A/T			
rs3888943	25537	34806887	A/G			
rs3888944	25554	34806904	A/C			
rs132618	27963	34809313	A/T	0.44	0.40	0.330
rs132619	28134	34809484	G/T			
rs3827346	28356	34809706	A/G	0.84	0.85	0.568
rs132620	29648	34810998	-/A	0.30	0.31	0.665
rs132621	29986	34811336	A/G	0.32	0.32	0.955
rs80575	30217	34811567	G/T	0.26	0.27	0.812
rs80576	30267	34811617	A/G	0.22	0.29	**0.024**
rs80577	30315	34811665	A/G	0.27	0.21	0.072
rs80578	30585	34811935	C/T	0.50	0.46	0.263
rs80579	30724	34812074	A/C	0.22	0.24	0.639
rs80580	30897	34812247	C/T	0.32	0.32	0.976
rs132622	30931	34812281	C/T	0.30	0.32	0.701
rs132623	31080	34812430	G/T	0.57	0.52	0.224
rs132624	31246	34812596	C/T	0.36	0.39	0.393
rs132625	31373	34812723	A/T			
rs132626	31463	34812813	A/G	0.86	0.84	0.723
rs132627	31467	34812817	A/G	0.14	0.12	0.672
rs1807672	32188	34813538	G/T	0.31	0.31	0.940
rs132628	32288	34813638	C/T	0.26	0.23	0.307
rs132629	32520	34813870	A/T	0.05	0.07	0.493
rs132630	32594	34813944	A/C	0.76	0.80	0.269
rs132631	32657	34814007	A/C	0.74	0.67	0.058
rs132632	32677	34814027	A/G	0.64	0.68	0.316
rs132633	32764	34814114	C/T	0.33	0.34	0.757

dbSNP rs#	Position in SEQ ID NO: 1	Chromosome Position	A1/A2 Allele	F A2 Case AF	F A2 Control AF	F p-Value
rs132634	32784	34814134	A/G	0.45	0.43	0.574
rs132635	32830	34814180	C/T	0.42	-0.01	
rs132636	32872	34814222	C/T	0.40	0.40	0.856
rs129603	33121	34814471	A/C			
rs132637	33348	34814698	G/T	0.09	0.09	0.882
rs3788518	33952	34815302	C/G	0.15	0.19	0.133
rs132638	34184	34815534	C/G	0.58	0.55	0.324
rs132639	34361	34815711	A/T	0.13	0.16	0.210
rs132640	35026	34816376	A/G	0.34	0.28	0.075
rs132641	35192	34816542	A/G	0.49	0.47	0.539
rs132642	35600	34816950	A/T	0.14	0.19	0.079
rs132643	36033	34817383	C/T	0.47	0.45	0.565
rs132644	36289	34817639	C/T	0.57	0.51	0.156
rs132645	38869	34820219	A/G	0.17	0.23	0.076
rs2017329	39629	34820979	A/T	0.39	0.42	0.531
rs739198	40530	34821880	C/T	0.64	0.05	
rs132647	41621	34822971	C/T	0.25	0.22	0.339
rs2097465	42379	34823729	C/T	0.54	0.50	0.311
rs2105915	42802	34824152	C/T	0.58	0.56	0.667
rs132648	42865	34824215	T/C			
rs132649	43644	34824994	A/G	0.21	0.19	0.471
rs132650	45051	34826401	C/T	0.31	0.34	0.365
rs132651	45828	34827178	A/C			
rs132652	45829	34827179	A/T			
rs80584	46257	34827607	C/T			
rs132653	47286	34828636	A/C	0.15	0.18	0.404
rs916334	47427	34828777	C/T	0.36	0.33	0.383
rs132654	47963	34829313	C/T	0.53	0.60	0.090
rs132655	48013	34829363	C/T	0.39	0.44	0.150
rs132656	48229	34829579	C/T	0.36	0.39	0.507
rs3834684	48282	34829632	-/A	0.21	0.21	0.956
rs132657	48376	34829726	-/G	0.51	0.45	0.112
rs916335	48404	34829754	A/G	0.36	0.40	0.245
rs132659	49900	34831250	C/T			
rs132660	52699	34834049	G/T	0.33	0.41	**0.044**
rs132661	52897	34834247	A/G	0.29	0.27	0.624
rs132662	53414	34834764	A/G	0.59	0.59	0.998
rs132663	53487	34834837	A/T	0.27	0.26	0.792
rs132664	54112	34835462	G/T	0.28	0.28	0.934
rs132667	55492	34836842	A/G	0.36	0.42	0.054
rs132670	59766	34841116	C/T			
rs132671	60307	34841657	A/G	0.50	0.49	0.839
rs132672	60701	34842051	A/G	0.21	0.25	0.297
rs132673	60952	34842302	A/G	0.42	0.40	0.525
rs132674	61401	34842751	C/T	0.33	0.32	0.909
rs132675	62379	34843729	C/T	0.38	0.36	0.722
rs80585	62870	34844220	C/T	0.27	0.30	0.405
rs80586	62879	34844229	A/G	0.68	0.62	0.113
rs132676	63499	34844849	A/T	0.25	0.27	0.442
rs132677	64284	34845634	-/A	0.69	0.64	0.147
rs132678	64408	34845758	A/G	0.43	0.45	0.620
rs132680	64760	34846110	A/G	0.19	0.23	0.258
rs132681	65230	34846580	A/G	0.23	0.28	0.201
rs132683	66127	34847477	A/G	0.19	0.21	0.493
rs2269594	66634	34847984	C/T	0.68	0.69	0.866
rs132684	66686	34848036	A/G	0.21	0.24	0.487
rs132685	66694	34848044	C/G	0.29	0.29	0.888

dbSNP rs#	Position in SEQ ID NO: 1	Chromosome Position	A1/A2 Allele	F A2 Case AF	F A2 Control AF	F p-Value
rs132686	67113	34848463	A/G	0.45	0.44	0.708
rs132687	67257	34848607	A/G	0.96	0.93	0.210
rs132688	67403	34848753	A/G	0.22	0.23	0.917
rs132689	67609	34848959	A/G	0.63	0.62	0.839
rs132690	68418	34849768	-/A	0.16	0.19	0.295
rs132691	68610	34849960	C/G	0.53	0.50	0.441
rs132692	69629	34850979	C/T	0.64	0.59	0.223
rs132693	70024	34851374	A/G	0.57	0.52	0.191
rs132694	70848	34852198	A/G	0.17	0.20	0.413
rs132695	71428	34852778	C/G	0.23	0.23	0.928
rs1966266	71553	34852903	C/T	0.49	0.49	0.881
rs1966267	71633	34852983	A/G	0.41	0.38	0.400
rs106808	71768	34853118	A/C	0.71	0.63	**0.030**
rs132696	71769	34853119	A/G			
rs2239829	73039	34854389	A/G	0.35	0.33	0.665
rs2285154	73325	34854675	A/G			
rs2239830	73412	34854762	A/C	0.51	-0.02	
rs2239831	73547	34854897	C/T	0.51	0.53	0.577
rs3865722	73769	34855119	A/T	0.56	0.57	0.887
rs3865723	73806	34855156	A/G	0.61	0.60	0.837
rs3985996	74467	34855817	C/T	0.30	0.27	0.501
rs3985997	74472	34855822	C/T	0.89	0.92	0.287
rs3985998	74473	34855823	A/G			
rs3985999	74482	34855832	C/T	0.18	0.17	0.691
rs3986000	74494	34855844	A/C			
rs2413382	74592	34855942	A/G	0.60	0.62	0.642
rs2413383	74670	34856020	G/T			
rs2413384	74672	34856022	G/T			
rs2413385	74714	34856064	G/T	0.70	0.72	0.433
rs2413386	74723	34856073	A/T	0.70	0.72	0.510
rs1894606	74749	34856099	A/G			
rs916336	74861	34856211	C/G			
rs916337	74892	34856242	C/T			
rs916338	74893	34856243	C/T	0.39	0.43	0.334
rs132697	75176	34856526	A/G	0.60	0.57	0.463
rs12781	75705	34857055	A/G			
rs1053983	75989	34857339	A/G	0.43	0.39	0.279
rs1053982	76027	34857377	A/G	0.74	0.74	0.852
rs2227167	77949	34859299	A/G	0.35	0.35	0.998
rs2227168	77974	34859324	C/T	0.38	0.42	0.226
rs132700	78167	34859517	C/T	0.25	0.25	0.989
rs3075364	78310	34859660	-/CT	0.41	0.37	0.252
rs2227169	78415	34859765	C/T	0.36	0.38	0.520
rs2097466	78575	34859925	C/T	0.52	0.57	0.255
rs2097467	78590	34859940	C/T			
rs2413387	78709	34860059	C/T			
rs132701	78875	34860225	C/T	0.23	0.27	0.255
rs132702	79864	34861214	C/T			
rs132703	81316	34862666	C/T			
rs2269595	81320	34862670	A/G	0.18	0.19	0.672
rs2269596	81409	34862759	C/T	0.39	0.43	0.258
rs132704	81737	34863087	C/T	0.67	0.64	0.391
rs2007468	81843	34863193	A/G	0.41	0.37	0.247
rs132705	82102	34863452	C/T	0.32	0.33	0.785
rs2007706	82833	34864183	C/T	0.36	0.41	0.226
rs132706	83461	34864811	C/T			
rs132707	83624	34864974	C/T	0.27	0.26	0.598

dbSNP rs#	Position in SEQ ID NO: 1	Chromosome Position	A1/A2 Allele	F A2 Case AF	F A2 Control AF	F p-Value
rs132708	83660	34865010	C/G	0.31	0.30	0.754
rs132709	83701	34865051	A/T	0.43	0.45	0.627
rs132710	83708	34865058	A/G	0.63	0.60	0.482
rs132711	83782	34865132	C/T			
rs132712	85707	34867057	A/G	0.87	0.86	0.667
rs132713	85717	34867067	A/G	0.28	0.24	0.293
rs132714	86486	34867836	C/T			
rs132716	86833	34868183	A/G			
rs132717	87115	34868465	C/T	0.47	0.49	0.533
rs132718	87234	34868584	A/G	0.68	0.66	0.705
rs132719	87479	34868829	G/T	0.62	0.57	0.223
rs132720	87561	34868911	A/G	0.12	0.11	0.727
rs132721	87604	34868954	A/G	0.75	0.77	0.495
rs132722	87674	34869024	C/T	0.21	0.20	0.630
rs132723	87958	34869308	A/G			
rs132724	87992	34869342	-/G			
rs132725	88019	34869369	A/G	0.69	0.00	
rs132726	88074	34869424	C/G	0.12	0.10	0.509
rs132727	88079	34869429	C/G	0.68	0.70	0.583
rs132728	88115	34869465	A/G			
rs132729	88118	34869468	C/G			
rs132730	88120	34869470	A/G			
rs132731	88135	34869485	-/CTCAT			
rs132732	88142	34869492	G/T			
rs132733	88143	34869493	G/T			
rs140575	88149	34869499	ACA/TG	0.30	0.29	0.903
rs132734	88340	34869690	A/G	0.47	0.49	0.571
rs132735	88344	34869694	G/T	0.88	0.90	0.595
rs80587	88512	34869862	C/G	0.37	0.33	0.365
rs132736	88521	34869871	C/T			
rs132737	88650	34870000	C/G	0.31	0.29	0.686
rs132738	88827	34870177	C/T	0.82	0.81	0.773
rs1807673	89230	34870580	A/G	0.26	0.23	0.356
rs2014700	89236	34870586	A/G	0.93	0.96	0.324
rs132739	90754	34872104	G/A			
rs1812023	90984	34872334	A/G	0.71	-0.01	
rs1812024	91110	34872460	A/G	0.70	0.70	0.888
rs2005590	92026	34873376	C/T	0.69	0.70	0.719
rs132740	92954	34874304	C/T	0.08	0.09	0.442
rs3986001	93375	34874725	-/TTGC	0.53	0.55	0.563
rs2413390	93794	34875144	C/T	0.40	0.43	0.365
rs132743	94937	34876287	C/G	0.07	0.09	0.436
rs132744	95068	34876418	C/T	0.74	untyped	
rs2413391	96188	34877538	A/G	0.43	0.41	0.611
rs132749	97092	34878442	C/T	0.47	0.48	0.858
rs132750	98812	34880162	C/T	0.70	0.67	0.285
rs132741	not mapped	not mapped	A/C	0.30	0.30	0.999
rs2413388	not mapped	not mapped	A/T	0.30	0.30	0.989
rs2413389	not mapped	not mapped	C/G	0.28	0.26	0.555

[0231] Allelotyping results were considered particularly significant with a calculated p-value of less than or equal to 0.05 for allelotype results. These values are indicated in bold. The allelotyping p-values were plotted in Figure 1 for the discovery cohort. The position of each SNP on the chromosome is presented on the x-axis. The y-axis provides the negative logarithm (base 10) of the p-value comparing

the estimated allele in the case group to that of the control group. The minor allele frequency of the control group for each SNP designated by an X or other symbol on the graph in Figure 1 can be determined by consulting Table 13. For example, the left-most X on the left graph is at position 34781551. By proceeding down the Table from top to bottom and across the graphs from left to right the allele frequency associated with each symbol shown can be determined.

[0232] To aid the interpretation, multiple lines have been added to the graph. The broken horizontal lines are drawn at two common significance levels, 0.05 and 0.01. The vertical broken lines are drawn every 20kb to assist in the interpretation of distances between SNPs. Two other lines are drawn to expose linear trends in the association of SNPs to the disease. The light gray line (or generally bottom-most curve) is a nonlinear smoother through the data points on the graph using a local polynomial regression method (W.S. Cleveland, E. Grosse and W.M. Shyu (1992) Local regression models. Chapter 8 of Statistical Models in S eds J.M. Chambers and T.J. Hastie, Wadsworth & Brooks/Cole.). The black line provides a local test for excess statistical significance to identify regions of association. This was created by use of a 10kb sliding window with 1kb step sizes. Within each window, a chi-square goodness of fit test was applied to compare the proportion of SNPs that were significant at a test wise level of 0.01, to the proportion that would be expected by chance alone (0.05 for the methods used here). Resulting p-values that were less than 10^{-8} were truncated at that value.

[0233] The exons and introns of the genes in the covered region are plotted below each graph at the appropriate chromosomal positions. The gene boundary is indicated by the broken horizontal line. The exon positions are shown as thick, unbroken bars. An arrow is place at the 3' end of each gene to show the direction of transcription.

Example 5

In Vitro Production of Target Polypeptides

[0234] cDNA is cloned into a pIVEX 2.3-MCS vector (Roche Biochem) using a directional cloning method. A cDNA insert is prepared using PCR with forward and reverse primers having 5' restriction site tags (in frame) and 5-6 additional nucleotides in addition to 3' gene-specific portions, the latter of which is typically about twenty to about twenty-five base pairs in length. A Sal I restriction site is introduced by the forward primer and a Sma I restriction site is introduced by the reverse primer. The ends of PCR products are cut with the corresponding restriction enzymes (*i.e.*, Sal I and Sma I) and the products are gel-purified. The pIVEX 2.3-MCS vector is linearized using the same restriction enzymes, and the fragment with the correct sized fragment is isolated by gel-purification. Purified PCR product is ligated into the linearized pIVEX 2.3-MCS vector and *E. coli* cells transformed for plasmid amplification. The newly constructed expression vector is verified by restriction mapping and used for protein production.

[0235] *E. coli* lysate is reconstituted with 0.25 ml of Reconstitution Buffer, the Reaction Mix is reconstituted with 0.8 ml of Reconstitution Buffer; the Feeding Mix is reconstituted with 10.5 ml of Reconstitution Buffer; and the Energy Mix is reconstituted with 0.6 ml of Reconstitution Buffer. 0.5 ml of the Energy Mix was added to the Feeding Mix to obtain the Feeding Solution. 0.75 ml of Reaction Mix, 50 µl of Energy Mix, and 10 µg of the template DNA is added to the *E. coli* lysate.

[0236] Using the reaction device (Roche Biochem), 1 ml of the Reaction Solution is loaded into the reaction compartment. The reaction device is turned upside-down and 10 ml of the Feeding Solution is loaded into the feeding compartment. All lids are closed and the reaction device is loaded into the RTS500 instrument. The instrument is run at 30°C for 24 hours with a stir bar speed of 150 rpm. The pIVEX 2.3 MCS vector includes a nucleotide sequence that encodes six consecutive histidine amino acids on the C-terminal end of the target polypeptide for the purpose of protein purification. Target polypeptide is purified by contacting the contents of reaction device with resin modified with Ni^{2+} ions. Target polypeptide is eluted from the resin with a solution containing free Ni^{2+} ions.

<u>Example 6</u>

<u>Cellular Production of Target Polypeptides</u>

[0237] Nucleic acids are cloned into DNA plasmids having phage recombination cites and target polypeptides are expressed therefrom in a variety of host cells. Alpha phage genomic DNA contains short sequences known as attP sites, and *E. coli* genomic DNA contains unique, short sequences known as attB sites. These regions share homology, allowing for integration of phage DNA into *E. coli* via directional, site-specific recombination using the phage protein Int and the *E. coli* protein IHF. Integration produces two new att sites, L and R, which flank the inserted prophage DNA. Phage excision from *E. coli* genomic DNA can also be accomplished using these two proteins with the addition of a second phage protein, Xis. DNA vectors have been produced where the integration/excision process is modified to allow for the directional integration or excision of a target DNA fragment into a backbone vector in a rapid *in vitro* reaction (Gateway™ Technology (Invitrogen, Inc.)).

[0238] A first step is to transfer the nucleic acid insert into a shuttle vector that contains attL sites surrounding the negative selection gene, ccdB (*e.g.* pENTER vector, Invitrogen, Inc.). This transfer process is accomplished by digesting the nucleic acid from a DNA vector used for sequencing, and to ligate it into the multicloning site of the shuttle vector, which will place it between the two attL sites while removing the negative selection gene ccdB. A second method is to amplify the nucleic acid by the polymerase chain reaction (PCR) with primers containing attB sites. The amplified fragment then is integrated into the shuttle vector using Int and IHF. A third method is to utilize a topoisomerase-mediated process, in which the nucleic acid is amplified via PCR using gene-specific primers with the 5'

upstream primer containing an additional CACC sequence (*e.g.*, TOPO® expression kit (Invitrogen, Inc.)). In conjunction with Topoisomerase I, the PCR amplified fragment can be cloned into the shuttle vector via the attL sites in the correct orientation.

[0239] Once the nucleic acid is transferred into the shuttle vector, it can be cloned into an expression vector having attR sites. Several vectors containing attR sites for expression of target polypeptide as a native polypeptide, N-fusion polypeptide, and C-fusion polypeptides are commercially available (*e.g.*, pDEST (Invitrogen, Inc.)), and any vector can be converted into an expression vector for receiving a nucleic acid from the shuttle vector by introducing an insert having an attR site flanked by an antibiotic resistant gene for selection using the standard methods described above. Transfer of the nucleic acid from the shuttle vector is accomplished by directional recombination using Int, IHF, and Xis (LR clonase). Then the desired sequence can be transferred to an expression vector by carrying out a one hour incubation at room temperature with Int, IHF, and Xis, a ten minute incubation at 37°C with proteinase K, transforming bacteria and allowing expression for one hour, and then plating on selective media. Generally, 90% cloning efficiency is achieved by this method. Examples of expression vectors are pDEST 14 bacterial expression vector with att7 promoter, pDEST 15 bacterial expression vector with a T7 promoter and a N-terminal GST tag, pDEST 17 bacterial vector with a T7 promoter and a N-terminal polyhistidine affinity tag, and pDEST 12.2 mammalian expression vector with a CMV promoter and neo resistance gene. These expression vectors or others like them are transformed or transfected into cells for expression of the target polypeptide or polypeptide variants. These expression vectors are often transfected, for example, into murine-transformed a adipocyte cell line 3T3-L1, (ATCC), human embryonic kidney cell line 293, and rat cardiomyocyte cell line H9C2.

Nucleotide and Amino Acid Sequence Embodiments

[0240] Table A includes information pertaining to the incident polymorphic variant associated with osteoarthritis identified herein. Public information pertaining to the polymorphism and the genomic sequence that includes the polymorphism are indicated. The genomic sequences identified in Table A may be accessed at the http address www.ncbi.nih.gov/entrez/query.fcgi, for example, by using the publicly available SNP reference number (*e.g.*, rs132659). The chromosome position refers to the position of the SNP within NCBI's Genome Build 34, which may be accessed at the following http address: www.ncbi.nlm.nih.gov/mapview/map_search.cgi?chr=hum_chr.inf&query=. The "Contig Position" provided in Table A corresponds to a nucleotide position set forth in the contig sequence (see "Contig Accession No."), and designates the polymorphic site corresponding to the SNP reference number. The sequence containing the polymorphisms also may be referenced by the "Nucleotide Accession No." set forth in Table A. The "Sequence Identification" corresponds to cDNA sequence that

encodes associated target polypeptides (*e.g.*, APOL3) of the invention. The position of the SNP within the cDNA sequence is provided in the "Sequence Position" column of Table A. If the SNP falls within an exon, the corresponding amino acid position (and amino acid change, if applicable) is provided as well. Also, the allelic variation at the polymorphic site and the allelic variant identified as associated with osteoarthritis is specified in Table A. All nucleotide and polypeptide sequences referenced and accessed by the parameters set forth in Table A are incorporated herein by reference.

Table A

RS_ID	Chrom-osome	Chrom Position	Contig Accession No. [1]	Contig Posi-tion	Nucleotide Acces-sion No. [2]	Sequence Position	Amino Acid Posi-tion	Locus	Locus ID	A [3]	Allelic Variability	OA Assoc. Allele
rs132659	22	34831250	Hs22_11677_34:9	15897265	NM_014349	mrna-utr		APOL3	80833	F	[C/T]	C

[1] Contig Accession Number which can be found in the NCBI Database:
http address: www.ncbi.nih.gov/entrez/query.fcgi

[2] Sequence Identification or Nucleotide Accession Number which can be found in the NCBI Database:
http address: www.ncbi.nih.gov/entrez/query.fcgi

[3] "A" column is the sequence orientation ("F" is forward, "R" is reverse).

[0241] The following is a genomic nucleotide sequence for an *APOL3* region. The following nucleotide representations are used throughout: "A" or "a" is adenosine, adenine, or adenylic acid; "C" or "c" is cytidine, cytosine, or cytidylic acid; "G" or "g" is guanosine, guanine, or guanylic acid; "T" or "t" is thymidine, thymine, or thymidylic acid; and "I" or "i" is inosine, hypoxanthine, or inosinic acid. Exons are indicated in italicized lower case type, introns are depicted in normal text lower case type, and polymorphic sites are depicted in bold upper case type. SNPs are designated by the following convention: "R" represents A or G, "M" represents A or C; "W" represents A or T; "Y" represents C or T; "S" represents C or G; "K" represents G or T; "V" represents A, C or G; "H" represents A, C, or T; "D" represents A, G, or T; "B" represents C, G, or T; and "N" represents A, G, C, or T.

APOL3 Genomic Sequence (SEQ ID NO: 1)

```
>22:34781351-34880400

1       tacacttttt tttttttttt tgagacagag tcttgctcta ttgcccaggc tggagtgcag
61      tggcactatc tcagctgact gcaacgcctc ctgggttcaa gctattttca tacctcagcc
121     tcccaagtag ctgggattac aggtgtgtgc caccatgccc agctaatttt cgttttttaa
181     gtagagacag ggttccacca Ygttggctag gctggtctca aactcctgac ctcaagtgat
241     cctcctgcct tggcctccga aagtgctggg attacaggcg tgagccacca aacccggcct
301     gatacacaat tgtctatggg agaagcacag tgaatgcact ctgatttagc aatcttattc
361     tcttgaatgg tttgtgcttt ttatgtctta ggaaaccttc cttggtccca agatattcta
421     tagaRacttt attttgtact tttcttctaa aagtcttttc tttttttttt ttttttgaga
```

481 cagagtctca ctccgttgcc caggctggag tgcagtggca cgatcttggc tcactgcaac
541 ctccgcctcc tcggttcaag cgattctcct gcctcagcct ccagagtagc tgggactaca
601 ggcatgcgcc accacgcctg gctaattatt gtatttttag tagagacggg gtttcaccat
661 attggccagg ctagtcttga actcctgagc tcaagtgatc tgcccatctc ggcctcccaa
721 agtgcaggga ttacaggcgt gagccaccac gcccagcctc taaaagtctt aacgttttgc
781 ctttcacatt tacatctgga attgattttg ttatgcggtg aggaaagaat ctttttattt
841 ttctctttgg ggactcaccc cccactccat ttgcaggcct gttgggaggt aggctcatct
901 tagaggaaca gaccctgggc aggctgtgac acagggctcc agcaacaagc cgcagaagcg
961 gaactcataa catgtggcat tggcattccc taccattact gattcaatgc cattccacgg
1021 tggccctggg tcatggacca gctcatgttg ggctaatgtc ccaacacaca caaagatggt
1081 aggtttgact aagcMagagg cagtccttag gtgaggttgc agggagaggg aagagggctt
1141 ggtgaagaat agaagaccca tgaagaaggg gtgtggtctt ggtgaccacg tgatcccata
1201 gtattcccat agacactcaa gacagctggt aaattgagct gctccccaga ggaggccaac
1261 agacaggctg gacaatgact ggggtctgag cagagacccc gtctcccttc tgctgcacct
1321 gctgcagctt aacattgact agactaccag gctcagcctc agcaaggccc ccacctggct
1381 ctgggcctgg tgaccaggga aggtgggaag agaggggggct catcacttcc tacagaagga
1441 agctcagact tccaatttcg tgtggcataa gagcatgggc tctgggcaac actgcctgga
1501 ttagtattca gctcgggccc tactcgtggt ataactggga gtcactaatt ttatctctgg
1561 gcttcactat aaatgggatc ctaatggcag tctacccaag cagggtattg tgaggcctag
1621 agatgatgag gcaaacagtt ggttctcaat acaagaaaga aatttatcat ttgcatgtag
1681 ctgcatagct gataagccca tgacagtttt cccataacct gttttcatgg ccttattttc
1741 cgcaactctg tctatttcag ctacgtggaa ccactcctgt gtactgtgac agtctaatgt
1801 tgggatctgg ttaaatatat tatggcccat ccaggcgcgg tggctcacgc ctgtaacccc
1861 agcactttga gaggccaaga caggcgaatc accctgaggt caggagttcg agaccagctt
1921 ggccaagatg gtgaaacccc gtctctacta aaaatacaaa aattagctga gtgtggtggt
1981 gcatgcctgt aatcccagct actcgggagg ctgaggcagg agaattgctt tgaaaccggg
2041 aggtggaggt tgcattgagc caagattgca ccactgcact ccaacttggg cgacagacag
2101 agagagactc ggtctcaaaa aaaaaaaaaa aaaaagttat ggcccatcca caggatgcaa
2161 tatcatatag ccatttaaaa gaatgggcta catctgaacc Mgatggaaga aaagatgccc
2221 ggatatgtta ttaatgggat gaggcaacgt gtagaacctg tgtaaaatag gatttttttt
2281 ctgagatgga gtcttgctct gtcacccagg ctggagtgca ttggtgccat ctctcctcac
2341 tgcaacctct gcctcctggg ttcaagcgat tctcctgcct cagcctccca agtagctggg
2401 attataggag cccgccacca catctggcta gtttttgtat ttttagtaga gacgaggttt
2461 caccatgttg gccaggctgg tcttgaactc ctgatctcag gtgatccacc cacctcggcc
2521 tcccaaagtg ctgggattac aggtgtgagc caccacgcct ggccaagtag gatgctttta
2581 tgccatattt aaaagaatcc acagggagct gttaatgctg atacttttgg ggaaagggct
2641 cagtggagca gctaagcttt tgcttttctt tttgttcctt tctgtggaat tggcattgcc
2701 taccaacact gattcagtgt cattccacgg tggccctggg gcatggtgtt ttttttcatg
2761 gatcagctct gcctgacacc accatgcagc aggctgtact gtgagcctca ttttgcaggt
2821 gagagaagca aggcacagag aggttacgta acttgcttaa agacatgtga caagtaagag
2881 gtggagctgg gattcaaact caggcagtgg aggtctggag tccctatttt tagctactac
2941 atgacactgt ctcccaatat attattttta tctatttaaa aattcaaagc agttttctgt
3001 gttatagact attttttact tcttcacatg cttaaaaaca aacaaacaaa aaaaaataaa
3061 acagaggcca ggcacggtgg ctcacgcctg taatcccagc actttggaag gccgaggaag
3121 gtggattacc tgaggtcagg agttcgagac cagcctggcc aacatggtga aaccctgtct
3181 ttactaaaaa tacaaaaaat agccaagtgt ggtggcatgc gcctgtagtc ccagctactc
3241 gggaggctga agcaggagaa tcgcttgaac ctgggagaca gagtttgcag tgagctgaga
3301 ttgtgtcatt gtactccagc ctgggcgaca gagagacttc atctcaaaaa aacaaaacaa
3361 aacaaaacaa aacaaaaaaa aacaggaaca atgcatagcc catgatgaat tgttttatta
3421 tctgtgactc ttattgttca ttggttgctt ctgtccctct ctgaccacag cctgatggga
3481 ggcacaggcg agggaagaac gtggctttgg gagcaggggg accctcattc tcatcccagc
3541 tctgcacgaa ctgcggcatg accttggcca ggccccttcc cttccttggc ctcagcctcc
3601 atctgtagag ctgggccaat gccggagcca atgacctctg cgggccctcc tgcacaagga
3661 ttctctccct ctttctgatc tatttttgcc tgacgcctct cccaggggcc tgtgggcctg
3721 agctaatttt cctggctggc cccagtccag gcagctcagg tggagagcgc caagctaatg
3781 aggccaaggt catggtccag ttagcatctc tcggaggaac cgtggctttt cagccagact
3841 gtagctccct gaccagtcac tgcgctcagg tagaggcggg ccctagagga gaggacaata
3901 gcaaagctgg atcctgcaag tcagcatttc cagtaggagc tgctcagggc ccagggccca
3961 gataccсttt gtgactgcag ggcagcatgt ccaggctcac atcaacatac caggatgatt
4021 caaggagtgt gtacgttggg ggttagggag aggaccaggt ggatcaccaa gtgcctgcaa
4081 gtacccccac acctgtcaag tcaggaagat gagaaccaat tctaccttta aatctaaagc
4141 agattgggag gccgaggcag gaggatcacc tgaggtcagg agtttgagac cagcgtgacc
4201 aacatggtga aaccccatct ctactaaaaa tataaaaatt agccaggcat ggtggcgggc
4261 gccacctccc agctactcag gaggctgtgg cgtgagaatc gcttgaacct gggaggcgaa

```
4321    ggttgcattg agccaagatc atgccattgc actccagcct gggggataga gcgagactct
4381    gtctcaaaaa acaaaacaaa acaaaacaaa caatctatat ctatatctat atctatctat
4441    ctatctatct atctatctat ctatctatct atctatataa agcagacagc atgtccacat
4501    tctctcaagg tgaccccctc tttctaacat cagaactttc cttttttcta catgaggcat
4561    ttccaccact tttaaatcag atgtaagaca gagtaagcct tcctctttta cctgccacga
4621    gggccagggg ttgtctgttc cttctctact acatccccgg cacctagacc catgcctggc
4681    gtacctacac taggagctca gtcaatattg ttgagcgaat caaatcacca ctgttccaga
4741    atgaggctgc agactgcctt tgaacggaca cagtattgcc agtgggtttg actcagtttt
4801    gttcaaatat ttcaatttct ttttttggtt tgtttttctg agggagtctc actctgttgc
4861    ccaggctggc gtgcagtagc gtgatctcgg tcggctcact gcaacctcca cctcctgggt
4921    tcaggcgatt ctcctgcctc agcctcccaa gtagctgggg ctacaagcat gcgccaccac
4981    acctggctaa tatttttgta tttttagtag atacagggtt tcaccatgtt ggtcaggctg
5041    gtctcaaact actgacctca aatgatccac ctgcctcagt ctccgaaagt gctgggatca
5101    caggcgtgag tcaccgtgcc tggcctcaga tgcttctatt tcaacgagtc tggtgaagct
5161    tgttatttag agagtaatat gtgaatgctc tagaaaaagc aattctcctg tctcagcctc
5221    ccaagtagct ggcactacag gtgcccacca ccatgcctgg ttaagttttt gatttttact
5281    agagacaagg ttttaccatg ttggccaggc tggtctcgaa ctcctgactt caagtgatct
5341    accctccttg gcctcccaaa gtgctgggat tagaggaatg agccaccaca cttggccaaa
5401    ttattccttt tacaatctgc tacaaccccc tgatttgaac attttgcaaa tttgaatgtt
5461    ctcattaggt tttcgattca gttcagtgaa tgtttattat taagtattgt tattatatta
5521    ttattcatta ttgttattat ttaataagaa taattgagtt cttaccataa gccagacttt
5581    gggttaaaca cataacatgt agatttagtt aattttcatg acaaatctat aagatgggca
5641    ctactgctat cctactttac agatgagaaa actgaggcag acagaggtta agtaaattgc
5701    tcaaccccac atgaatagta agttctggag ccaggagtca aacttcagag cctacactcc
5761    taatcacagc taatgttgct cttcaggacg tactttacat gcaagtacag taactgtcta
5821    caaggcactg gtgattaaaa aatgaacaaa gcctttaaca ggtgaatggt taagcagact
5881    gtggcacatc cacatcatgg aacacaattc agcaacagca cggaatgaac actattgatg
5941    cacacagtaa cctgaatgac tctccagaga attatgccaa ctgcgaaaag gcaatccgca
6001    agagttacgt actctatgat gccatttact tatttatttt gagacctagt ctcgctctgt
6061    cacccaagct ggagtgcagt ggcatgatct cggctcactg taacctccgc ctcctgggtg
6121    caagcgattt tcctgcctca gctttccaag tagctgggac tacaggcgtg caccaccacg
6181    cccagctaat ttttgtattt ttagtagaga cagggtttta ccatgttggc cgggctggtc
6241    tcgaacttct gacctcaaat gatccgccct ccttggcctc ccaaagtgct gggattacag
6301    gcgtgagcca ccgtgcctgg aaaattattt atttttaaga tttgttaagt gtagtcatga
6361    gaagcgggaa agagtagaac aaggagtttt atctgctgtg actgaacaat caattgagat
6421    aacgcggcac cttcggatca gctgattcca tttatctaac attcttaaat ggacccactt
6481    atagacatga agaatagatt agcaattgcc aaggctaaag tagggatcag ggtgggagag
6541    aagtgggtga ggctataaaa ggacaacaag agggattctt gtggtgatag aaatattcta
6601    tatcttgatt atatcaatgc caatatcctg gttgtgatat tgaagaggct aaagctgcat
6661    ttggtacaat aaaatacaat tagcagccta taaccccaaa ggagaaaact ggccagttgt
6721    gtattaaatc tttgtgtgtt tgcatatgtg acatttaaat gcttaattta tatgccaaca
6781    tttagcattt ttccaagatg ttactattaa cggaaactga gtaaagggta catgggatct
6841    ctctggatta ctttgtacaa atgcaaataa aaatgttaat ttaaaaaaac agaatgaggc
6901    caggtacggt ggttcatgcc tgtaatccca gagctttggg agaccaaggt gggaggatcg
6961    attgaggcca ggagttcaag accagcctgg ccaaccagca agaccccatc tctacaaaaa
7021    attaaaaaat tcagctgggc atggtggcgt atgcctgtag tcctagctac tgcagaggct
7081    gaggcaggaa aatcacttga gactgagagg tcgaggctgc agtgagcaat gatggtgcca
7141    ctatactcta gcctgggcaa cagggcaaga gaatgaggca gacaggagcc cagcctggca
7201    ggaaagggag gttgtattag tcagggttct ccagagagac agagctaata ggatagatgt
7261    atatatgcag gggagtttat taaggagtat tgactcacac gaacacaagg tgaagtccca
7321    caacaggctg ttttcaagct gattagcaaa gaagccagcc agagtctcaa aacctcaaaa
7381    gtagggaagc tgacagcgca gcttttagtc tgtggccgga ggcctgagag cccctggcaa
7441    accactggtg taagtccaag agtccaaaag ctgaagaact tggagtctga tgttcaaggg
7501    caggaagcat ccagcacggg agaaagacca tggaagggag tgaaggctgg aagactcagc
7561    aagtctagtc cttccaagtt cttctgcctg cttaattcta gccgtgttgg cagctgatta
7621    gatggtgccc acccagattg agggtgcgtc tgcctctccc aggccactga ttcaaatgtt
7681    aatctccttt ggccacaccc tcacagacac acccaggaac aatactttgc atccttcaat
7741    ccaatcaagt tgacactcaa tattaactat cacagagggc ttccaatgca gtgacagttg
7801    acatgagtcc tcaaggttag tttgcaggtc acaagggaga gaaggacagg aggcctgagg
7861    caggcaggac agcatgtgRt agtgggctgt ttgggtcttc cttccccaaa gtcatcctgg
7921    gatggtgaga tctccggcca ttcattccat tgcacagagg cattcatgca ggctacctcc
7981    caacttgttc ataaaaatgt taagtgatgg acaaaataat gtctttacat ttttcttaaa
8041    cgtttggttc ctgctgctgc ttatctggta gagaggctca tcagaggaga aaatgaagga
8101    agattccctt gtcaaacatc acaacagaga gcaagcatct ttgttcagcc tgtcaggatc
```

8161 atctttcaca ttcaatttag aaaaatggca gatgttaatg tttttcctgt taagagataa
8221 tttcttttta aaatgtttgt tttttttttc cccttactat gcctagtcct gctttttata
8281 tgataacagc ctaggagtgg aaagaatgag agtttgggtg acagagatcc ttacatttga
8341 atctcaattc cacttagagg ctatggtcat gacttagagg ctatattaag accaYttaga
8401 ggctatggtc atgaccttga gtccgtttcc ctgagtttca ctgcctcaga ctaatcatac
8461 Ygaataggcc agtattgcca aatcccagtg ctaggatggg gattaatgag aaaacatttg
8521 ccaaaaaacc tagtatacct tgtgtaatac ctgacacgta gtaggtgctc aataagtgat
8581 actcattgat ttattttttt attttattta ttttattttt ttgagaccga gtcttactct
8641 gtcacccagg ctagagtgca gaggcgtgat ttcgacttac tgcaacctct acctcctggg
8701 ttcaaatgat tctcctgcct gggcctcccg agtagctggg ataacaggcg tgtgccacca
8761 cacctggcta atttttatat ttttagtaaa gacagggttt caccatgttg gccaggctgg
8821 tcttgaactc ctgacctcag gtgatccgcc tgcctcagcc tcccaaagtg ctgagattac
8881 agttgtgagc caccacgcct ggtcaataag tgatactcat tgttaataac aattctagtt
8941 catctaagga tttaaagctg tgtacctttc atagtgttct acaaacagct tcctgatttc
9001 cagcagcttt gcttatatca gccaggatag tctaggttat gctgcagtaa aaaacaagcc
9061 cctaggctgg gcacggtggc tcacgcctgt aatcccagca ctttgggagg ccgaggcggg
9121 cggatcacct gaggtcagga gtttgagacc agcctgacca acgtggtgaa cccctgtctc
9181 tattaaaaat tcaaaattag ccgggcatgg tggtgcatgc ctgtagtcct agctagttgg
9241 gaggctgagg caggagaatc gcttgaacct gggaggtgga gcttgcagtg agccaagatc
9301 gcgccattgc actccagcct gggctactag agcgaaactc tgtctcaaaa aacaaaacaa
9361 aactaaacaa aacaaaaaaa caaaccaaac aaacaaacaa aaaccaccaa aaagcaagcc
9421 cccgatctca gtggttcaca acagcaaagg ctacttcttg ctcacgctac atgtccatca
9481 tggggcaggt gcctctctgc tcYggttgtc ttccctcctg gactcagtat catggagcaa
9541 ccactttctg gaagattgct ggttgccaca gtagaggaaa agggtcttat atcagcattc
9601 attgctgtgg cttagaaatg acacatatta tttccattca aaactctttt gtccagaact
9661 ggttacatat tcccacccaa ccataaggag accaggaaga ataatcccac tgtgcatctg
9721 gaaaggggaa aagccagaca aatgtggcca acagccccag tgactaccac cacagtgctt
9781 ttgaaaacca ctgcccagag ccagaaggat ggaggatcag catggcggta ggaggaaggg
9841 agaagccaga ggggaatgag gaggagccgg ggtgcctggt atggagggcc atccatcctt
9901 ggtcaatgtg cagggctggg gccagaggaa ggctgggcgc ctttctcatc tggacccaga
9961 cccagacaga ccccggggcc attccgttct gcaggcaaac aagccagtcc cgctagaatg
10021 tggcagtcac aggctttcaa catggaaggt ggagctgcct ggacgctgtg ttttttttct
10081 gtttgttttt ttgagacgga gtcttgctct gtcacccagg ctggagtgca gtggcatgat
10141 ctgggctgat tgcgcctcgg caaaatctcg gctgacttgc ctcccgggtt caagcgattc
10201 tcctgcttca gcctcccgag tagttgagat tacaggcgcg caccaccacg cccggctaat
10261 ttctgtattt ttagtacaga cggggtttca tcatgttggc cacKctggtc ctgacctcaa
10321 gcgatccgcc cgcccaggcc tcccaaagtg ctgggattac aggaataagc caccgcgccc
10381 ggccagacgc tgtgttttcc tggcactcca cttcatgagc aaaactgtca cagagtagta
10441 tttccctggc tccacccctc cggcgctgct ggattcctga acctattcac ttcttctgaa
10501 gtgctctcaa cttggcctcc ctggctcctt cctggtgcag acatttactc ttcctgctcc
10561 ggctttgtgc ctctgtaaaa ctgctgacgt caccagcagc ggctccgacc gcctgtcctc
10621 caaggtcccg ggctgcggct gcgcagcgcc gacagcacga tgcgtgtgcc gacagcgcca
10681 cctcctggct ctgcYgaggc tgttggcaac acgcagggct cgcagtgcgc ccgcttttct
10741 actgcagccc tcaaaattca gcacgaggct tggctagggg agcttgagtg gctgacactg
10801 tggtctggtt agtttccatt ctgacccatg ctagccgtct ttcagattct agaactcagt
10861 tcttctccgg acctccctcc cagccacaag tctttgaata tgcggcctgc agctgggctc
10921 ccctcatccc cgctcttctc cccaaccttc ttcatcatcc cttggccgat cttgaagact
10981 tagcggaggc agggcgcaat ggcttacgcc tgtaatccca gcacttaggg aggccgaggc
11041 gggagcatca cttgaggtca ggagtttgag accagtctgg ccaacatggt ggaactttgt
11101 ctctactaaa aatacaaaaa ttagctgggc gtggtggcgg gcacctgtac tcccagctac
11161 tggggaggtt gaggtaggag aatcgcttgt acccgggagg cagaggttgc tgtgagccag
11221 gattgagcca ctgcactcca gcctgggcaa cagagcgaga tcctgtctca aaggaaaaaa
11281 aaagtcttag ctgaaacgtc acttcctcga ctcctcaaat ttagtcaagc ccccgtgcta
11341 tgagatctct ccccttcttg ctggacgggt tggcaatgct ggaggatctt gaaagaggaa
11401 cactcatggg tcaagagcat ttcctaacag cgaacagagg tttcagaatc ctcatgagca
11461 tcggtgtctc cttgctggga cgctgctcac acactttgtg atcctcagag gtgagatgca
11521 ggcctattaa acacgctggc cttggtgaga aagcctctgt gggtttcctg tttatcccaa
11581 taagctgatt caggaaagga gacatctcct tccttctctg tttgccatgg aggactcaca
11641 cctttcttgt ccccaggaca gtggacatct gtttgccaat gcagacctcg tgctcactgg
11701 aggctggaga aggcacgctc aagaaacgtt caaggcatga cagctaccca gaaggcctct
11761 agtgccccaa gagctgctca tcagccccat ttcttcctta atattattta aaaatttaaa
11821 gaaattcagt gaaagtgtaa gcatacaggc ataggtagta ggtaaactag aattcactct
11881 tgctgggggga ataccaccca gggtttgaaa agatacagaa ggcaaaagaa actgcattta
11941 gtataataaa atacaattag gcagagcacg gtggctcatg cctgtaattc cagcactttg

12001 ggaggccgag gaaggcggat tacctgagct ctggagtttg agaccagcct gggcaacatg
12061 gtgaaacccc gtctctacaa acatacaaca attagtgagt cttggtggta catgccagtc
12121 atcccagtta tttgctgggc tgggctggga tgggaggact gtttgggctc aggaggtcaa
12181 ggctgcaatg agctgtgttc acgccactgc actccagcct gggtgacaaa gcaagaccat
12241 gtctagaaaa aataaaaaaa ttagcagcct ataattccaa agcagaaaac cagctggctg
12301 gctatgcact aaatctgacc agccattgtg tgtaaggcgt gtgtgtgcaa atgtgtgttt
12361 gcatgtatat cgtttaaata gttactttat ttaccaacat ttacatttaa ggaaatttca
12421 catcaaataa tctataattc tagaatcttc ttaaacatga ggtgggcaac atgagacccc
12481 cacttctatc agtatccatc ctgactggac ccaagtggac agcttgagtt tccagttcaa
12541 tccagttcct gcctggcctc tgctggcctc tgggtttgca atcccaactc cctaccaccg
12601 ccactactgc tgagagcatc ccctctacag tactcagaac ctcactggct gcctgcaacc
12661 ctgggggatc agcagaagag actctaagcc catctatcag gtaaaaaacc tgaggcccag
12721 agagggaaag tgacttacct agggacacac agctcttagt aaacagagtg gaggccaact
12781 ccactccaac tcagtctcat tcccactctg agatttaaga aaaaaaaatg tgattttaaa
12841 aatcccaaag ttgtttcctt cctccctccc tccctccctt cctccctctc cctcctttcc
12901 ttcctccctc tccctccttt ccttccttcc tctccctcct ttcctccctt cctctccctc
12961 ctttcctccc ttcctctccc tcctttcatt ccttcctctc cctccttttc ttccttcctc
13021 tccctccttt tcttccttcc tctccctcct tttcttcctt cctctccctc cttttcttcc
13081 ttcctctccc tccttttctt ccttcctctc cctccttttc ttccttcctc tccctccttt
13141 tcttccttcc tctccctcct ttcctccctt cctctccctc cttttcttcc ttcctctccc
13201 tccttttctt ccttcctctc cctcctttcc ttccttcctc tccctccttt ccttccttcc
13261 tctccctcct ttccttcctt ccttcctctc tccctccctc cctctctccc tccattcccc
13321 ttccttccct ccctcccttg ccctccctcc ctcccttctt tccttccttc ctttgtttct
13381 tccttccttc ttcccttcct ccctctctct ccttccttcc tctctttctc tctctctctc
13441 tctctttctt tttccccctc tctctctttc tttctttctt ttgacagggt ctcagtcctc
13501 actctgtcac caaggctaga gaacagcagt gtgatcatgg ctcactgtag cttctacctc
13561 ctcgggttga agggagcctc ccacctcagc ttccccagta gctggaacca cagtcacgag
13621 ccacaatgcc tggctaattt tttgtatttt ttgtagagat ggggtttcac gttgttgccc
13681 aggctggtct caaactcctg ggctcaagca gtcctcttgc cttggcctcc ggaagtgctg
13741 ggattacagc catgagccac cgcgcccggc caatcgcaaa gtttctaaag taggaaatgg
13801 gattctattc catccccaac taattgtccc caaccactgg tagcagcctg gcttgtctct
13861 gcttcgacaa acacagaggt ctaagctcac actctcaaag ttacactcac atgtctacag
13921 tttggctttc tctcttttac tttttttaaa atttattttc tatttgtttt ccccatgaca
13981 tagctccagg agatcctgag aacatgtccc ctacagtttg gcttttaatc ttatgattgc
14041 aaagctggga tcatgctgca aacaccactc tgcaacgtat ccttcccacc aacccagcgc
14101 ccacacaacc ccctcagtgt gggccatctt tccctttcag gacacacagc acagtccctt
14161 tcttccctca tcacctcaca ggcctgaaca gagggccaga gaggtggagt cctgcaattt
14221 acctgctcag tctcccatca caggaccctg gacaagtcac tcctctctca gagcctcagt
14281 ttccctgcag gaaagaggag gtgctcagag acaaggtgga tgaaacaatt tctgggcttc
14341 ttgtatctca aaagaaatcc tcttcctgtg gccgggcgcg gtggctcttg cctttaatcc
14401 cagcactttg ggaggctgag gcaggcggat cacctgaggt caggagatcc ccgtcaggag
14461 atccatgacc atcatggaga aaccccgtct ctactaaaaa tacaaaatta gctgggcatg
14521 gtggtgcatg cctgtaatcc cagctactca ggaggctgag gcaggagaat cgcttgaacc
14581 tggcaagcgg aggttgtggt gagccgagat cgcgccattg cactccagcc taggcaacaa
14641 gagcaaaact ctgtctcaaa aaaaaaaaaa aaaatcctct tcccctgtca ggccttagta
14701 gcagcatcct aggggtagcc tagacggggg cagattcctg gggtggttc taacccacat
14761 cacattaatg atgccaccca agtccattcc catggcacca ggtcactggg acaaagggcc
14821 aggtctccag tcttccatgt ctgacctcta cctcctttgc ttcctccaaa gtgcccagag
14881 aaatgcacgt ggcccacggc tgctgccttc tgaggtcacc tggccatacc ccaagtagtc
14941 tgggaaccca tgctgagcct ggagaatctt cactctccca cctctccctc cattcctcag
15001 gggaacaagg actccaaaag agccgtcgga taaacctggg tttggatcca ggctttctac
15061 cttggacaag ttacctaaac ctcctgcatc tcggtttctt cttcgtcaaa tggggacaat
15121 acttctttct ttaaaggctt tactttgagg atgaggtgag agggaagcca tcctcctccg
15181 taccccacaa ggatcacagg ccagtgagat ccagctatta tcacctctgg tgatgctggt
15241 cgccaacaac cccaaggaccc ctgttactga gctctaacca gaccctccct ctagtatcca
15301 acacatactc tcaagggaaa cagaaatcac acagaagagg ccaggcaggc gcagtggctc
15361 acgcctgtaa tcttagcatt ttgggaggca gaggcgggtg gatcacctga ggtcaggagt
15421 ttgagaccag cctggccaac atggtgaaac cctgtctcta ctaaaagcac aaaattagcc
15481 gggcgtggtg gcacacactt gtaatcccag ctactcagga ggctgaggca ggagaatcgc
15541 ttgaacccgg gaggcaaagg ttgcagtgag ccaagatcgt gccactgcac tcaagcctgg
15601 gcaacaagag cgaaactctg tctcaaaaaa aaaaaaaaaa gaaagaaatc acatagagga
15661 aataaattgg ctggttctat tctcaacctc agctccgctg ctccactggc taactcagag
15721 gacagcccga gtccagccat gacccactgt ccagcctcct ttcctgcagg gaaccagctt
15781 gggtttccag gacgacacac tggcccgaag tcttactccc agtcgccact gcactcctgc

15841 ctgggcaaca acagtgaaac tccatctcaa aaaaaaaaaa aaaaaaaaga atgatacaac
15901 ggactctggg aactcacggg gaaggttgga aaggccaggt atgcgggatg gtattggcaa
15961 ggacacatat agagacacag agaaggagca tcgtaggggg tatggggaag ctatgcacat
16021 ttcaaagtat tagtagagca caaggaagag gctgggacag ggggaggatg aacttggaga
16081 agccatgact ggaagcctgt gtcccatggc attgtcatct gtgtgtgtgc cccatcttga
16141 ctttgagctc cacacaggac aggactctga ctcatctttg ttatagatct cagtgagtgc
16201 aaggcacgaa gaagaagggg tgtcaatgag tgcttcataa atgaatgatt ggatgaatga
16261 atgaatcatc tgtttctcca cttccattct gtggtctccR gccaacattc cctcccattt
16321 acctctcatt aaactctttt tgcagcagca ctttctctta gttcttcctg tcatttgcta
16381 gacttcagct ttgtgatctt tgggtcagac accatgatga gagatgcaga tcgcaactga
16441 gcgKcattta aatattcaga agagctgagg ctccgcaggc tgaattctga atatgcagat
16501 aattcagaaa acactggagc aatagacttg ttgaattcaa actcgaggtg gagcagacga
16561 gggtacagct cggcttaggt gacccatcgt ccccatcagc tgctccttcc caaagctacg
16621 gctccctcgg cctttgagct ttcaagcctc agcctcgtgt gttgcttcag tggcagctgg
16681 aacattcctc cctgcagcct gctcctcaga cgcttatctc accccatagg gccctgctcc
16741 atctcagtct tggatggggt tagcagtgtt ttctggatca aagtagcttc tttccttgtg
16801 ctgtatttaa actttgtggc gtaggcatca gtttccctaa tgcctctgtt atccatcttg
16861 caggtctaca gtattgagag acagaaagtc aagaagttct ggacctgcca ggaattccca
16921 ctgtggaaac aaatccaccc tccttacttc ctggagaggg gcaattcctt caggaaggtc
16981 ttccatgagc cttacatcag actttgtcct gaactccacc cacagctccc ctgtctgagc
17041 gtgtcccatt ttctctggac ctctttcatc tccagtccca gaccagcatc ctcctgcatg
17101 tgcttctgaa agtacaggcc actcctgcat cccttctata cctgggtcgc caggaaactc
17161 acggatccac ttttgacgca gtctgttctg ctctcccttc tcaaatctgc ccagatctca
17221 gctgtccccc atggtcctta tcaaactccc atctacagcc tctagaaagt gacttctgat
17281 catttcctct atcgtggccc tgaatggcag tcacatgtac ccttcatctg tttggctctt
17341 gccccctgcc agacccagca gcaattccat gcccctgtct ttctggatct cagctcgaga
17401 ggagtttgag gtgcgtggct tgattttcgc ccatcattct gtcctgtctg gcttgcatgc
17461 cttacctttt cttcacagga aatcccatca ttcaatccac atgttgcttt tccatgtacc
17521 aatattagat tttcctagtg attcttttct gatagtaagt cccaaatctg tcttttagcc
17581 aatcttatcc Ytccctgatc ctctgagaaa ccagactcct agatacggac ctttcaaatg
17641 cttctttgct tctgggcttt taataatccc cctgagatta ttcaccagct aaaggatggt
17701 ggggacagaa cagcacttga ctgataaagg cccccatcct cctggtctgt cactctcaca
17761 tcccaccgtt ccctggggta ccccctggag gccagcacac gacctctcca gatccctgtt
17821 gacaccactt gttggggaaa ggcagctcga gcatcccatg gacatccgtc tgaaaggttg
17881 gccccaatct ggttccttcc ttgatggagg acgacattaa aagaagacaa aggagagaat
17941 ttggggtgac tgatctattg ctgcttaggc caagtctaac cattgaaNgt ctccaaatct
18001 cacacctaag ggtccctagc tcttgtagaa caccacagag attttccagt tcaaggtttt
18061 ggtctatcac acgatgggag aaagctggca agaagaaaac aagctgtggg aatcaacaat
18121 tgttttactc accaaagatt aggaagcctg aataacaaca tcacctcttt tgagtacagc
18181 ctcacataga actgaactga atagaaatca tctcatttta caatcatcag acatctgtct
18241 ttgccccagt cctggctgct gcaccctcac tacgagccct ggattgccac gtcgtccctt
18301 ggttcccaaa tcctagcccg tggatataca tttcttatcc taagagtcac cctactatgt
18361 gagggttctg ggagcagctt tgttcctggg aatctgcagg caccttcctg ccttcctgct
18421 gtggtcgcca ggcacaagga aagaaagaag ggtctcgctt attacagtag ggaaaacatg
18481 cgatgaattc tcttctaacc ctggtgcaga aattctccta atgagtcctg ggtgaaacg
18541 caggcctgga agcgaggtga ggctcctccc caccagcatc taaaaagccg gctccctgga
18601 gccacaggat ttttccaatc agccatggga ccagggcagc ccaccagcct ccccagctgc
18661 cagacgctgg tagagagacc tgagacagag ggcaggtgaa tggagcggga gaggaggaga
18721 gaagggacac cagatggggt ctgaggaaga aagaaggaag aaaaaaagat accttgatgg
18781 ggaaggggaa agccttaatc tgcttctcaa tcgcgaagta ggtctgcttc tcaatcatgc
18841 agtagggagt gaaaagcaca ggtggggctt ttgaggctca ggcactccac ccaccaggac
18901 aatgcggacc ccatgactgc agagcctagg ctctgcaggg agcagctagg gagacactag
18961 taagggaggt gggcaggggg cagaggcccc tttcccccac tgctcccacg gggatgagga
19021 gggggcttcc gaaaaggccc aggttagggg gcctgcctgc tcagctcaca tccccctggt
19081 gtgacagtgc cttctttact ctgtcctcta gccccNggga cccatctgag tctctcccct
19141 ggaaattgga ggacctgaaa gaaaaccctt ggaggcagca ggcagtggag ctgaggcctt
19201 gagtgtagct cgttccctac cgctgagagc cctgaggctt tctgggtccc tggtcttccc
19261 taagcctgtt tttaaaaaat gtgcacgcac gcccacccac ccaaggattt ctttaattct
19321 gtgagctttc tctctggatc attcaaatat atgcctcYac tcttgaggta gcgggaagat
19381 tttttaagat ttttcatttg tttttgtttt tgcatttgct tttgtcttgg ttttgctttg
19441 ctatgttcta cttgcatcca aagggtccta cctaatagaa aaattggcac caaagggttg
19501 gttgcagacc acaaagcctc aggataattg gtacaggtag aatcatgttt tgtgttgaga
19561 ggtggacaga gcaatagtga acccttctca cccccctcat tctggggagc acctgctctc
19621 agcttgtgga aaggggacaa ggaatgctgg gacatgaagc aggaccactc actttggaca

19681 aaggcctttg acatttctcc tgaaggcgtc caataaattg tggctgttga tatttttggc
19741 tgactctcct atgaaaaaca aattcagaaa ccgtttctat aggtttctgg ttacatgatg
19801 aatttcatcc tcactagatc ttctaagcgc acatgagggc cctgatgggg acagagtggt
19861 gtgctagtac ctctcctgct gtaacaaagc cctacagaca cagtgacttg aaccaacaca
19921 aagttgttct cttccagtag agttctaaaa tcaaggcatg ggtagagctg cattccttgg
19981 ggaggctctg ggagaagcaa ttcccttacc ttttccagct tttagaggca cctgctgtcc
20041 tcagctcctg gcttctccct tcgtcttcaa gccagccttg tagcatcctc tgatctctct
20101 cagtttctct ccccttctat catctcatct ccttttctcc ctctcaccct actgtctccg
20161 ccttataagg accccttgtgg ataccttggg cccacccaga taatcgcaga taacctctca
20221 tctccagatc cttgactgga tcacacctgt gaggtccctt ttatgatgga aggtgacata
20281 atcacaagtc ctatgattaR gacacccaca tctgtcaggg gctgttattt agcctaacac
20341 aagatgtatt cagaaaattg caatagggat acctgagggg aattgaagga tcccaactcc
20401 aagccgcatc taaggatgct tccgaggtgc agctcccgct ggaggactct gctccatctc
20461 tcttcctgtc tgccgctgcc tgcctctctg tttctagaca gccagtgttc aagtcagtat
20521 ggccttgaga gtgaactgca gaatggaaag gagtgagatg ctgcagcagg agaaagagct
20581 cgctgtcaaa gctccgaaga gctgaagacc atttatttgc atggtggggt ctgacggagg
20641 tcggcaagga agaggaatca tcatttggWt tccttgccat ggtcatcctc agctggcacg
20701 gtttgctcca gcagcaggct gggattccag ctgtgcccag tgttagccat gccaccctca
20761 actaaatgct gatcacttct ttgtgacata aaaagccaga aattagttgg catgatgaag
20821 aagaaggagg actcaaagtg tgtaggaggc tgaaatgcta gatttgattt ctcaagcact
20881 tcctgcccac Rctgtttcct gccagggccc aaaagatttc ccgctccttt aacgagaaga
20941 gtcaatggat aaccaggaac tctctataat aggggagggg ctgggactct ggtctcttta
21001 gggtgggaga aattccaaga ggccgactgt aggggaggca ggtaactgcc agaagaaaag
21061 agagtgcggg cgccaggtca cgcaccagga gggcgcatga tcatggctca agggacattt
21121 gataggaaca aggttccaaa atagagctgg aagacagtcc actgacttta caaaaccccc
21181 aatctggtga gtggatacct gtcacaggtc atcaggctgg atagtccctg cttgtcaccc
21241 aattcttagg cctgggtcag agggggagat tatagggcac aagtaattag tgaatatcta
21301 tattcgttta cctttgatta agtccatcag ggtcccaccc tcacctggtc atcatttttc
21361 aaattcctgg gtggataact ggaaagaata ctcttggaag cttcagattg ttcacccaga
21421 cccacggacg gggactattt cagtgacata Ygccagtgga ggccattggt tttacatacc
21481 tgcctaaata ttttactcaa agtcaacacc aacttcatgg aggagttaca gaggcagtta
21541 cagagatttc tgccatcatc aagatttaaa gggagcagaa gcagtgattc ctgtgactca
21601 cctgttcacc tctccagtgt ggccagtcgt caacagggaa acaactgttt gttttccaac
21661 cgcagaacaa agtaggaagc tcatcaatgg aagttgccat tcctaagttg ttcttgttaa
21721 tgaagcaaat aaacacagcc tgcagctctc agtctgtatc aatcaggctt cttggtcgag
21781 cagaggtacc aactctgttt tcttagggca gaaaacaaca tgaaattcac tggaatgata
21841 ccagaagaac cccacgtgtg taggaggctg gaaattggct tggaagataa atagaagcca
21901 agacagtgag acggtatcga aggctgcgag cacaagagcc ctgctttgtg aggagtgaca
21961 cggccacagc gctgcccYgc cacctgcgga cgtcatctct gctggtgcag gactcggtcg
22021 tccccgtctc tattcccctc ccgggcagtc ctgctgctct cactgagaat ggattctaag
22081 caccctgggc tttgcctcac ttgctccaga ttcaggatcc catgtggaag cagccaatta
22141 gccagcctag gtcttgtcca gaagccctgg atgccagggt gggggtggac agagagtcat
22201 ctattctctt gggcattaca tagaaggtgg gaccagtctt tctctatgac tccacatgct
22261 gggaaacccc acaaaactag gaagaggcga atagatgcca atcagctccc tcccaggaaa
22321 ctgggagttc attccaccca gaggcctaaa atgttgacat gccgtctttg agtgtagagg
22381 gaaggaagct aactccaatt cgcaaaatga ctaaatagag ataaatgact gacagctcag
22441 aaaagaccat gagggctgaa ttcaggtgct tacacagagt catgagtact cagcatctct
22501 caaattgtgc tgtgctttcc gcctggccca cctgctcccc ccaggtctgg cctgctgtgg
22561 aacgcctctc aaggacggtt cctgaagagc aatggagcac acctcacagt ccaggctgcc
22621 tggaaggagc tgcatcctcg gaccaggatc tgcattgttt cctgggtaac aactaaggaa
22681 ttagctggaa tttcaacatt ttagaaggaa taaatctgga ggcttggtga caaagatttt
22741 tagggaagag gaacatacat agagctttca gaacaacggt agaaStaaaa aaaaatctgt
22801 cccatgtgaa tgctcattgg aagatgtcca ccatcaaaaa agactctttt tttttttttt
22861 ttgagactgg gtctcactgt atcgcccagg ctggagtgca gtggtgcaat ctcggctcac
22921 tgcaacctct gcctcccagg ttcaagcaat tctcgtgcct cagcctccaa agtagctggg
22981 attacaggca cccgtcacca tgtctggtgg ctagtttttg tatttctagt agaaataggg
23041 tttcaccatg ttgaccaggc tggtctccaa ctcctgactt caactgatcc actcgccttg
23101 gtctcccaat gtgctgggat tacaggcgtg agtcaccatg cctggctcag aggctgttaa
23161 taaacaaggg aagaagtcac tctttaaggc tcaaatctcc tctgttcccc aggcagcctg
23221 acattttcag aatgggctca ttaaccaact tgcctaatgt ggagctctat gtaaccacac
23281 aacacggacc ccactcatca aggctgacca cttggcaccc agtgaatctc aaccccggaa
23341 atgacagcct taggtcccat ggaaatggcc caaaggggac tcagctccat atcagaggct
23401 gggagctgcc cttaggatgt gatatgtttt gaatccatac ccagaatttt ccccatttcc
23461 tgtaatagcc agaattcatg gagctgggca gtcagaggtt aagatgaggg tgatagctct

23521 atggttaccc acaataaacc actttcaaaa attttgttta tgatacatac tcctaactca
23581 cagactttta ggtctggaag tttcagtatc taaggttgac aaaagttgtt tcagcagaaa
23641 ccatagctgg cttctcatgc tccctgatgt agggatagca atgacgtcac tgggttggat
23701 ataatcatgc atcatgttcc tccagggaag ctcatgttgt tattacacag caggacagga
23761 ggaattgagg gatcccctag gattgctcct tgtaatactc taaagtgtta gagtattagt
23821 gtaaatgtga ataggaaatt tcaccaactt actgaggcca gaccatcaag gagcttagac
23881 accacgagaa tgactttgcc cataagtaag aggaacattg agtggaacaa ttcaagactt
23941 ctgatggtgg cagcaatggt cagtagtggc tgtagccttg accgtacctc tttattttct
24001 acaaatatac acacacaaaa attctgagac agaagctgcc aatggaccaa aagaaaaatt
24061 gatcgtgtgc tgaggccctc ctcctgtgac ataacttcag catgcagaag gtggagatgc
24121 tgtgtgtggc tctcactggg gaagaggcaa gtgcatttgc ctttttcctt gggatggctg
24181 ggttcatagc agataaggag agttcttttt gtttgtatga gaagaagagt gtgtgtgcag
24241 tagcaagYga ttgactgtat acaatgagca caaattcagg tggctgtttg gccagaggct
24301 tcccattagg gtttgtgtct ctgttaccca tccccattcc tgataccata gcctcactga
24361 ttagttcaag aagagtcact aacccaagct attctgtcta ctgagtatgt gaaactttaa
24421 tcagagacac taggaatcaa agtgtgtgga gctaagccag gtgcaggcag agccctgaaa
24481 gggcaaacca tatattcctg atgttggtaa tggtctgctc ctgtccttcc aaattctgta
24541 gattttgttc ttccttccat cctctgacct gctctagttt tcatccaatg aatcttctat
24601 tttgtttaag acagatagtt tctgtttctg ttgcttgcaa ttgaaaaatt tttaaacaat
24661 actagcagta cagtggaatt gcatcaaatg caggtagcct gatggcattt caactctaag
24721 tgtcagggca aatgctggct ggatcactga agaagtaaat tactcagtgY cagaaacatt
24781 caaagcttaa gtgaaatgta tgaaaagaat gggaattatc caggaattga tgaacaattg
24841 acaRacaaat gaaacacaat gttctggcgt taggcaagtc ggtggccgtt gctctgttga
24901 cttcttcagg aagaaataga tgaatcaact gaaggtcagg gaagccttgg ctgatggaga
24961 gttacaggga ctgaggcaag tgttctgcag gatgctacat aaaaaaggac taagtaggcc
25021 ctgggtgcac tttacgagtg acttaggaga ggttctgggg tgattKtgtg cattctgtat
25081 ctcttgcccc tgctcYgggc tccagctgag gcgcccagca gttctatatt ctatgcatgg
25141 aaatacagat aattaatgtc cactggggca aggccatggg ggacaggagc agtggttaca
25201 ggcctccctc tctcctccct gtagaggaac attctgaaga cgctccttga agactcctca
25261 ggaggtctcc agtcgcccac atgggtgacc ctatcaagaa tttctcccRg catcagcctt
25321 ctgtccttct ctgtgtcact cctMccagta ccccagtcct gctcctagaa tcacttccca
25381 aacaaatcct ctgcctggga gccttgccta gggctctccc ttcaggggaa ctcaggccac
25441 agcagcctac attagacagc atgaaggctc tgggcttttg tgctaagagt aataggaagc
25501 cattgaagac atggaagcat gggcttgaWg aaatctRtgt tcttacgggg caaMttggcc
25561 cagtgtggag gatgggcagg gatagtggca agggctgtat aacagggaag accattctgg
25621 aagcttttgc aggagtctgg gtgaaatgag agggctgact ggatgagggt gtttgcagac
25681 tggatgaggg tgtttgcaga ctggatgagg gtgtttgctg tctggctatc tctctctgtt
25741 cctctatttc tcatggactt tctctcaacc tacacacagg cccctctgat tttagtcagc
25801 cttgtgtgac ccatgtcatt cgtgggccct ggctgctaga tggggaccat gcctgcactt
25861 gggccaagcc agatgtctac aagctttccc tgagtgcctc acaggccact gggcactggg
25921 gtcgcccaga aatgaccgac cagaagaaca agcagaaggg gccatggagg cctagctccc
25981 tgagtccctg gaagaaggat catgaggaag tgaagtcagg caccgagagc ccaggggagc
26041 ccatccatca gcaccacaaa ggcacaggca tagatcagtg ttgcttgaac ccagatctcc
26101 agaacacaag aggagagagc caggtggacc tgagatatca tcttgtccag cccctctttc
26161 ccagctgatg aggaaactga ggctcagaga ggcggggtct cttgactgca ggctggttca
26221 tctcctcaat gcaccacaaa gttccttgat tggttcctgg catctggttt tgcttttctc
26281 ttccttgttt tttttggggg gggggacgcg ggggatggag tctcgctctg tctcccagac
26341 tggagtgcat tggcgccatc tcggctcact gcaacctcct cctcttgagt tcaagcaatc
26401 cgcctgcctc agcctcctta gtagctaaga ttacaggcac ccaccaccat gcccagctaa
26461 tttttgtatt tttagtagag aaggggtttc actgtgttgt cctggctggt ctccaactac
26521 tgacctcagg tgatccaccc gcctcggcct cccaaagtgc caggattaca ggcctgagcc
26581 accgtgccca gcctcctttt atttttcttc acaaatcaat gacttttcca atacagaatt
26641 tggagtttca ttggaatgtc atcgttccct cagtgtatcc tggccccaag cagggctgac
26701 atttgtcaag tcctcctgct gtctgctgac atcatcacca tcacctttac cactgttcag
26761 acaaaaactt cagacaggaa cttgctccca aagcgagaaa ggagccagga gaccaaggaa
26821 tgactcgaat aagtccagct tggtgagtag gtgagtttat taggacttac acacagggca
26881 ctcagcagga tggctctaga gatccggcct cccccagtct ctaaactgct tttcagttaa
26941 ttttctgtct ctttgcctgc tgtatatgag taatgagact gtttttcttg gtaggttctc
27001 gcatactctc caggatgttt gggtttttag agacacctgg tcctcagctg gggacaatgg
27061 ccatggctca ttacctggcc ttcagggttc aagcagggga catatacccc taaataacct
27121 aaaggggatc catcacacta caaccaccac ctccaccgcc atcatcaaga agccactggc
27181 tgactgagat acacttccag gaggacaaga cagagtggat gctggaaaga cagggcaggg
27241 gaccatcacc agggaaagac ttcattcttg gaaggacatc gaaccggggg caggtcggta
27301 gtggagccgc tgtttcttct gctgtatcca aaagttctaa ctcttcggct ttctgcattt

```
27361   tcagctcttt cttttcctgg ccttctcatt gctggttcct gcacacctcc cctctattcc
27421   tccccccaat atattcgtta gtctaaagga aatttcttct tcctattccc cacactctcc
27481   agtcccctct cctcccttat tccaggctcc agcattcctg ccttctcctt ggtgcacttg
27541   ccatctgcat taacccctcc cctgctgtgc tcagctacag aaatgccaaa gtcactaaca
27601   ctcagctcca taaactttac cttgccctct ttatccccct aataaaatgc ctgcattttg
27661   tcgtggcctg tgtatggctc acctcccctg ctggctgcac tggtctgggg tcagtgggta
27721   tggcatggat tcagacgctg atagatctga gtgagctcca ttagattctc ctccagctcc
27781   tgagcctgcc gcctcagctc ctcagcagat gcagactttg ccccctcatg caagtgcttt
27841   gactcgtata caaggttgac cacatccagt gcaaggaaga tgcctgaagt ggtcgcactc
27901   aggatccggg ctcctctgct cactgcccgg gtggtgcctg caatcgttct ctctgcttga
27961   ccWccacttc cagctgagat tcgccaggtg gtcacaggga gtcgggccct ggctctggct
28021   tgcctgatgg cacggatttc actcccaatg gtttgtgtgg cttcgtaata attattaaga
28081   agggaaagta agttgggtgt gatgtcacgc ataacttcct taaatacctt caaKcggtca
28141   atgctggttg cagtcagcct gctggcttca gcttctgctg atgatgtgta tgagtgctcc
28201   acgatgctgg tggtgatccc agtcacagca gacgctgctc ccagccctac cccagctgca
28261   gtaagggcca gactcgtccc tgctgtaaat ggtgccaaaa caagaccagc aagggacatg
28321   atgccagagg cagcgccagt ggagctggac accacRttgg agatggtgca gcctctgtgg
28381   acctcttcaa taccatttgc aagggcacga agcttttcta tggactcctg gatcttcctc
28441   ttgacttggg gaaactcttt caaaaaccat tccctaaact gctcatcttt ctgctgcaca
28501   tattcgtcct caatagctgc atatgttcta agcttcttca gagcttcgta gagagcatct
28561   gcctcatccc tgtaccaata aaggacagat gattaagaaa ggcagcttac ttatctgtaa
28621   aatcagctca ataagatctg ttctacataa atcacagagc tattactatg agtcaaatgg
28681   aaataaagct ttaaatttta aatggaaaaa tttaaaataa tttttccaag gttaaacatg
28741   ttttgtatat catagatatg ccacagtact tattctgtat gggaataaaa aacttaaagt
28801   catcttcaga tgctctttag tggtaggtat tcgataaaga tggctcctca agcaagagag
28861   gtcaacatca tccctccaat agtgtgtgga gaggggaaat tttgtagatg aacaaagaga
28921   agtgggttcg gagcaaaaag accttcattc agaggatcat agaacagtga aggattaact
28981   ttattagaag agaggtcgtt cttttgtctt tgggccctgg gaagggatct ctggggccct
29041   ggaatgtcct gcctggtagg aacatctttg tttccctggt ggtttggcta cagcacagta
29101   tagcaatgtt atgggtgatg ggttttgggg ctattacgta tcttttctgc cttccaaagg
29161   aactgggaac taagggtgtt agacctcagg gaggggctgg ggactccatg tcagccatca
29221   ggacagcttg tgatccagcc ccagaaagag ctctgtacat tgaggcatag gttgagaaag
29281   actttctcct tggccaaact ccagtcaggc tgctctgagc cttttctcct ctaggcgctg
29341   accttccatg cccatcccgt gcttgctggg cctgtactgc cccagttcag caagaatccc
29401   ccaaggcagt tagggagggt cctcccccact tgctgtctga gcattcttgc tgtctgatca
29461   cggtcttcac ccgctcctta cctgccttta gttgagcagc ttgttaggcc agtttagcag
29521   aaacccccca ttctggtgtc tccagtcaga actggtccat ctcccactcc tcaccctggt
29581   gttggctgtg actccccctc tgtctttact atattgggag ctgggctcag tctccctccc
29641   ctgtaacNca atcctgaata aaggcttcct gatctaatgt gattcacagg aatttttcct
29701   taactgggtg agcttctctg gttggcaaca ccccatgtgt tttgtcccac attcacacag
29761   gagaacacag cgtcctgagg acaacgaagc ttcacacctg aaaccctccc agactctgcc
29821   ctgtgtgtct cttccttagg ctgattttaa tctgtgttct ttccctgtga taaaccacaa
29881   tcgtgcaata acagttatca gtgagttacg taagtctttg tagtgaatta tcaaacctga
29941   gtgggggctt aagaaacacc ttcaaatctg caactggtgt cagaaRgaag ggtgccttca
30001   gacattgcat tttgactaac tctgagtcct taagctatca aataaaacct cagactggaa
30061   gagatcacct tgcctggcat gttctctctt ccaatgccac atgccctttg atcttctgac
30121   tgaaggccac tgggtccagc cactgccccc tttgactgaa aagtgaaagg cctggctatt
30181   agtcctggct gcacatgaac tctcaagcca cttgacKgcc tcctgacctg gcacctccaa
30241   gcttcagggt ctcactgtca catgtcRcca gccaaactaa tgccttagca ggtgacagct
30301   tcacctgtgt gcatRccaat gccatgtgca atttcatgga acctcttatg gactcatgag
30361   actttgaaaa tcatctctat aaatcactaa tgctgaaaag acctgcattt ggacaagatg
30421   atagagaaag tgtttttctc tttggttggg gacttgtgag tctcctgaga acccttcaaa
30481   ttacccagtt ggcagacact tgttccaact ctccaagctg gcccgaagaa acacttcctc
30541   tactctatgt agaacagaag agggtcccaa atggcagttc agaaYgcccc agttcccagg
30601   catgctgtca gggaaacagt cacccatggc ccactatgcc catcagtgat gggatcaaat
30661   cagggaggga agaaatatcc ctctcccact cactggtcct ctgcaagtcc atggccacgt
30721   tacMtttcat gccctcagtg cctccctctt ctcaccctac tcccagcctt ggtgcccgag
30781   ccttcctgct gctcctccaa cccaggccca cccctgcccc atcctgaccc ctggctctgc
30841   ccctgaaacg tctctctggg tctttggatg gcatcatgct acttggttcc ccacctYtga
30901   cctgctgttc attctcaggc tttaaccaca Ygcccatctg attgaactac tattagcatt
30961   ggagagatgg caggtcccag tctccaagac ataaaacaac agccccagag ctgtgcaggg
31021   acagagccaa gatccacccc cagctctgtc caactccatt ccagcctctc tgtgttctgK
31081   gtcagcagga ggatcctgct gtgtttgtgc agttcctcag aggtggaaga aaagggcccc
31141   ggacagcatc tccaggaaca cacattgatc aggtcaagat tttctggct tttttcaaag
```

31201 aaagtcctgc tgggcatcca actctaagaa aaacacttgg tcgggYgcag cggctcatgc
31261 ctgtaatccc agcaccttgg gaggccgagg cgggaggatc acgaggtcag gagatcaaga
31321 ccatcctggc taacacgatg aaaccccgtc tgtactaaaa atacaaaaaa aaWtagccgg
31381 atgcggtggc cggagcctgt agtcccagct acttgggagg ctggggcagg agaatggcgt
31441 gaaaccggga ggcggaggtt gcRgtgRgcc gagatcatgc cactgcactc cagcctgggg
31501 gacaagcaag actctgtctc aaaaaaaaaa aaaaaagaaa aaaaaagaaa aaaaaagaaa
31561 aacacttttg gtcctggtct ctccccagca atctttttaa aaggtaacat taacggagca
31621 cctaccatgt gtcagtcatt atgctaaata ctttacatgc attagcttat ttttctcaca
31681 ataaagctca atggatgaag aaaggaaggc agaaagtagt caagtagctt gcaaagctcc
31741 catcaccagt agatggcaga actgaccctg cccttgtggg cttcctctcc cctcagcctg
31801 aggtcactca ctgccagcga aggacttgga ggaaatattc ctcctggact gctgagaggg
31861 acattctcaa gttcagcaga gggggctgcc tggaggaggt gtgcctgcca gagaaaacta
31921 gcccagggga gatcaggatg gcatagccgg ggcgccccat ggaggtaacc ccacggaggt
31981 tacctgggca attcagccgc agtcacgaat ctcttccagg cttcattgtt agtcagcagg
32041 atttgcagat gcactgggct gactctctcc cggaagtatt tggtggcctc ttcagtaaag
32101 cgtttctttt ctacttggaa acaaaaagca taagattgga agaaagtgtg ctacagccta
32161 aatggcattg agaataaggt ggttcgaKgt taatcctcct caaccagctg tcacatgggg
32221 tatttttgat ggaggcatca gtgctataga ctgagttgtg tgccctccaa attcatatac
32281 tgagaccYta aaccccagta actatatttg gagattgagc ctttaaggaa gtcattaatt
32341 taaagagagt cgtaagggta gggccctaat cctacaggac tggtgtctag ttaaaagaaa
32401 aaaaagaggc agaagatctc tctctctctc ctcaccccac cccgccctcg ctgtctctct
32461 tgccttccag ccctctctcc ctgtctctct ttctctctct ctctctctct ctcacacacW
32521 cacacacaca cacacacaca cacatacacc ccccacactg ggaaggccat gtgaggacac
32581 agccctgagg cagMctctgt aagccaggag cagagccctc cccagacatc aaccctgctg
32641 gcaccttcac cttggaMttt ttgcctccag aactgtRaga aaatacatca gttgttggag
32701 ccacccagcc tgtggcgttt tactaaggca tcctgagaag agtgattcac cagggaagtg
32761 ccaYggtgct ttgtggagga accRatctat ttcagctgag aatcaccaga aagtgagctt
32821 tccaccatgY tttctcccca tgtacgggaa atattccagt gatcgcttcc tYctgccatg
32881 tgcctattgt caaacgcttt acaccagtgt cttcaaaatc tgagtttcga cccatcagtg
32941 ggttatgaaa atagttcatt ttgtcaagag cagcatttta aaaagaaaag agcattgaaa
33001 atatcagagt gtattgaaca cagcaagagt gagtattctt ttgtaagata tttgaaatat
33061 atataatata tatatacata agggtgtata tatatatacc cttatgtaaa ttcctgggat
33121 Mtagatgtgt ttaggatggc agcattttca atctgtgtca atatgacaca gggtgcatat
33181 gctgtatcac tgaccactcc cacactggtc aacgctacaa atgtttcaac cttgtaatca
33241 aaggtacaaa tgtttcaaca gagaaaccaa tgaatattca ccttaaatgg gatacataat
33301 attatcagta caaagcatct tagtacacag caggattact gcccagaKga gttgccccaa
33361 aaatgtgtgg tttgcaggga tttgggaatt atggaactgc agagaaggga ctgggaagca
33421 gtatggattt atgtatgcac tgggacatga ggtaaaatcg tttcctgccg tggacacact
33481 tgcgaggcct cattctcatt tcacagaagg agaaattgag acccagacag cgaaagacac
33541 cacccaaccc agcaagcagg aagctcagga tttgactcaa ggccagcgtt cccttgccat
33601 tactttgaag actccatata ttttgaaaat gcctgtggtt cctgcattct ccactgtagt
33661 atacctgtgg gattttaaac attttattga aaaagaatat ttcaacccta caccattgaa
33721 taaactacca tgcacagagg attcctcctg ctgctcatgg gcgctctcag atgggcagtg
33781 gagggggttc tttaggcggg caggggagca agaccctcag cactgatccc tgtggccctt
33841 ctgtgttacc tctgtcctct gcaggctggt ctcaggggac atggctcccc actatcacct
33901 ctgaacccag ctggccgcgt tattaaacca cagatcatct tgtttcggag aSagagtttc
33961 ttccttgagc ctcccagcca agcccttctg aaggattccg ggttggctg gtaggtcccg
34021 gcttctaatc ttccctgccc tctcctcatc ttggacagga ctcagcagag cccagtttgc
34081 tgtccagggc ggctcctgag cagatgtgac ttggttgcca aggcaactca gtgaggggag
34141 aaaacaggcc ttagggaaat gaacacagtg ctgacttggg tgaSgggtat tttgctttta
34201 ttttttcctc cttcactgaa aggatgaagg aaaccagggt tggatggagg gaacctccag
34261 cactagatct gtgtatttat tgctggggtg ggagtaccag ggtgttaaag caaacttaaa
34321 atgaacacct gggcaaacaa aagcaaacag gccttcagaa Wggctgtaac ccccccttaaa
34381 ctgccaacta tggggaactt aactggagtc gtttcagatg ggtgcttacg ttaggcacaa
34441 acaaaactta acttcggcca gtcatgagca gccagctgac agacgggtac acgactagga
34501 attttccaac aaggtaaacc aaaaaacata atttgacaac tgcaacaaat caaataatgt
34561 ccttattcca cttccatatt caccctataa atacctgcct ctgacacttt tccatcataa
34621 catgaaatgt ctttcatttt gatgtttccc acttcatgac ttgcttcttc ctcaaataaa
34681 gtcttcaaaa ttcgattgga cctcagattt ttctttgaca aggggtattt gagatcatgg
34741 gaatgaggcc ctgcctgagg cagtcatccc tggtaagcag gtgcttgtga tgggccgtgc
34801 agggcaggcg ccctggggga caaaaagggg cttccacctg gaatcaacag tggcaggtaa
34861 aggggatggg aaccaactct tagggaggcc ctactgtgtg gcaggcactg gctgggccag
34921 gtacccatgt actatttgat cctcatgacc acatggggaa atattatact cgcattccag
34981 gtgaggggag gtaagaagct ccaggtgaca cagagaagac gtggcRaagg ccagaatggg

35041 gctcagttca gtgggacgct gagtcactct tttcacaggg ccactcagag cagaggggct
35101 ggtgcagggc aagagctgga gggaggggac tggcctgtgc tggaaacacc ccagggtcag
35161 ctgggagcag agcgggaggt cagaggtcgg gRctgctggt caacctcctc tctcagccct
35221 ccgcacctcc tggaccctgc tctcccccta ctccccgacc ctgctgtttc cacttcccca
35281 aggctgtggt gccccacttg agtcgtttgc taggtgccag ggcagggagg aagtgggcac
35341 ctgggccatg ctcccctgag gagaggtcag gacatagatg cccccacaga cccttgggtc
35401 gggggactcc atgtggcctc ttttggggct cagggaagac tcatcggcct gcttgaccat
35461 ccgggttcct ctggggctca ctgagttgtg gaaacgccac cctccattct aggtgcgagt
35521 aggaaccagc cagggaagag gaagaagcaa gcctacctga gtccatggtg ccagcagcca
35581 ggtcactgag agactttccW tggagctctc cagtcactga cctgactgga agggttcgtt
35641 ttttttctta acccttgagg aggagaaagc aaattgtggg actgaatgtg agccacctgt
35701 ggacagaggg aggttggagg ggctggatct ctgcaatccc tttgcgtgtc agcaaatgcc
35761 aagaccaacc tagcccggtg ttttttaatc tcctgggccc cagccagctg gtatttagag
35821 aaattctact tgctccagct ccctctgccc tcactctcac accaaggcag ggtccttttg
35881 tcctgtaggg gtatgaagag aagataatca gaggaaggga tggacatctg ataatcacag
35941 aaactacaga gtctatacac cgaaatagag atgggaagga acgttctgac aatgacctgg
36001 gcctggcaac atgtgtgata actaattgat aaYagccagt gtgtactggg tgtctgtgat
36061 ctgcagggct cctcatatgt gcctcattcg atgatcacac acaggccagg agatgggtgc
36121 aaccgtggct cccatttga aacagaagaa acctcaaatg ttaagcaacc ttgccatgtt
36181 ttcacaatag ggtagagcc aggacttctg attcaaatcc catgtcttcc atgactgggc
36241 tccactccct taaatatata aaaatattga ctcatgcata catgcttaYt gtaaatgact
36301 cctcattcca aatctgtgtc ttgcactcta tgcccaagtg tctgtcttgg tgaattcact
36361 ggtgctgtgc tggacaccct ttgtccccct tgcttctgcc tgtatgtcct ggctgcaagt
36421 gatgcttctg cttcaacttc ccaagtagtt gtccctacag gtacatacca ccacgctagg
36481 cacattaaaa aaaaattttt tttagaaatg tggatcttgc tttgttgccc agagctggtc
36541 tccaactcct ggtttcaagc aatcctcctg cctgagcccc ccaaagtcct gcgattgcag
36601 gcatcagcca ctacaaccag cctaaataaa tgctgatctt tatacagaga aacactcttc
36661 acattctgct tgtgatgatc ttgccttggg aaatgagaaa aaagaaagca aatgttgatt
36721 atgtttaggc attgttctaa gaaggtagaa gaacatttat tgactatccc ttcaagatct
36781 gactacagtg ttagtcaagg cacttatagg tataataaca gtcaatgaga aatagggtg
36841 aacatttgtg gactcgactt tgcaccacta attgctctaa gatgttataa ataataactc
36901 acagtcctct tccaagagcc tattaagtgg gtaccatttg tatttccatt ttcagataag
36961 aaaactgaaa cagatgttca gcgacttgcc caatctcaga aatctactaa ggtgtaaaga
37021 ctgggttccc acccaaccag acggactcaa gaacacacac actgcctctg caccctctgc
37081 tgccaatgaa aatgtgaatg gaccagttgt ggtggctcac gccttaatcc cagcaatttg
37141 ggagaccgag gcaggtggat cacgaggtca ggagatcaag accatgctgg ccaacatagt
37201 gaaaccctgt ctctactaaa aatgcacaca cacacacaca cacaaattag ctgggcgtgg
37261 tggcgtgtgc ctgtaatccc agttacccgg gaggctgagg caggagaatc gcttgaaccc
37321 aggaggcaga agttgcagtg agccgagatt gggtcactgc actccaccct ggtgacagac
37381 tgggactctg tctcaaaaaa tatatacata tgaaaatttg aatgaggaaa acaggagatc
37441 agagacttca acagatattt ggaagacaga actggctgaa gaagtgacaa ctgaatagag
37501 aggggtaaag ggatctgcaa tccacagcaa caccaagagc acattcagcc aagaattgcc
37561 tgctaagcag agcctcagag aggggagaag ttcaccagcc ccacgaccca tacaggattc
37621 acagtagccg tgagtctaca ccaatggaaa aaattagaga cttcataact gcatctaaag
37681 gccccagaaa taatatcttt gtatcctgat gccaaggagt tctcatttca tcactctaca
37741 gggaagactg ccagtggacc tgcccctttc tttaatatga atagatctag gatcagataa
37801 agcagaaaga caaaaaatgg ggaaaaggtc aagataacaa gtaaaaaatg agtccagaag
37861 aaatggatga ttgcgaaaat agaaaataat ttttaaaaac tcaaacccat gtcctcaggg
37921 acataaaaat gttttattca taaatgagaa cattatgtac tgaaaaggaa caaagatgat
37981 ggaacaaaaa agacttcttg agaattaaaa ggtgattgag aaaataaaca actctaggac
38041 aagattcaaa aagataggca aggaagcttt ccagaaaatc aactaaaagg atgacagaaa
38101 aagaaaaaaa caaaaacaaa aacagagcta gatatagaat aagaaagtta ccatacataa
38161 taattacata gaaagctact tgagtctgag catggtggct catgcctata accccatact
38221 ttgggagggt aagcaggaga attctttgag tccaggagtt ctagaccagt atgggccaca
38281 tagtgacacc ccatcattac aaagaataca aaaatgaacc aggtgtggtg atgcatgcct
38341 ctagtcccaa ctactctgga agctgactgg ggagaatctc ttgagcccaa gaggtcacag
38401 agcaagaccc tgtctctgtt aaaagaaaaa caagtaaata acagaaaaga aagttacttg
38461 agaatgagct gtggaggctg agaggaggaa gaatgtgagt gtgtgtgtgt gtgtgcacgc
38521 acactcatgg agaggggata aggaaaacca agagagaaaa agacgtgaaa gagaaaaatc
38581 aacccaggga gaaggaaatg caaggatcac agctgggcat gcagcaggcc tagaaagcaa
38641 ccagtccagg atagaataag aaggaaaggg ctctgtaggg aaaggggact cgatagaaca
38701 ttgcatataa ttcacagttt gagaaaaact gaggaaatca tgaaggcaaa cagtacaaga
38761 gggggaaaaa acaaaggcaa ttagaaacat caggaaaaac taaagccgtg cagcaaaaca
38821 tgtgcaacca tattgtacaa ctcagctcaa cagtaagcaa gatttaaaRc gtcaccaaaa

38881 tgtaaacact tggttaactt ttaactttta gaacaaaaca gaacatttac atatgagaac
38941 taaatctact tgttaccagt cagtataatt aatgttccat ggtaaaggtg ggaggtgaaa
39001 cactgaagga aggaggggga ggattaatat ggccattttg taaagtgaag agtcaagaaa
39061 tactggctac agttaagggc ataagaaata aagttaaata aactcacgtt tagttctcac
39121 aacaaccatc ctacgtaggc actgttatca tcttaatagt tctgagggga aaactaaagc
39181 ttggagaagt taagtaattc acctggagcc cctagctagt gacagccaga agccaaactc
39241 aagtgcccaa tctccagagc atgcactctc actctatttc atctgctagc ataagaatat
39301 tatgtaaaga aaatgattta ctccctaaag aaaagctggc taggcacact ggttcacatc
39361 tgtaatccca gcactttagg aagctgaggt gggtggatca cctgaggttg ggagttcgag
39421 agcagcctga ccaacatgga gcaactctgt ctctactaaa aaaatgaaaa aattagccgg
39481 gcatggtggt gcatgcctgt tatctcagct actcaggagg ctgaggcagg agaatcgctt
39541 gaacccagga ggtggagttt gtggtgagcc aagatagcta cattgcactc cagcctgggc
39601 aacaagagtg aaattctgtc tcagaaaaWa aaaaaaggaa aatctgaaca tcaaaattag
39661 ctgcttttaa ttgggaaggg agtgagtgga gccctcttta ttaaatgatt tgtctgaacc
39721 gtatacactt attaattgat ttttttaaat tttaacaaaa cagaagttac tttatagtag
39781 ttagtaactg ctaacctcaa taatggtttg gattgtggtg ccaccacctt aaagttgatg
39841 gagaagtgtg agcaagtgcc aaagagaaat acacctgcct cctgatggga gaaatgtcag
39901 cttcagagaa acctcggtgg ctggggtaac caagctgcta agaaaagtac ggatgttgtc
39961 atttatgtt aaactgttaa tattaacagc ataatcagag tgaactccta tcttgaatca
40021 gtataatttc cagtggcacc ttgccactca acatacgggc ttcttttcca cccaaaggaa
40081 ccaagacgtc ttgtggggaa ccaaattccc cagactgtga tatgtgttac cccagaaagc
40141 aaggacgctt ccaaagatgt tcagagcagt gttcaaaggg atgcccactg gtcagtccca
40201 accgctgtga acccggaaaa tctgagactg gtgtcagtta atttagaaag tttattttgc
40261 caaggttgag gacgtatgcc tgtgacacag cctcaggaag tcctgatgac atgtgcccaa
40321 ggtggttggg gtgtagcttg attttataca cttagggagc atgagacatc aattaagtac
40381 acttgagaaa tacatcggtt tggtctagaa aggcgggaca actcaaagtg ggcgtttcca
40441 ggctacaggt aaattgaaac gttttctggt tgacaattgg ttgagtttgt ctaaagaccg
40501 gagatcaata ggaaggaatg tttgagttgY gataagaggt tgtggagacc aaagttttat
40561 cctgcagatg aagctttcag ctagcaggct tcacagagaa acaggctgta aaatgtttct
40621 tattagactg aaagcctgtg ttgatattaa tgccagagag gcataatgaa gcatgtctga
40681 cccccatttc ccttcatggc ctgaaccagt ctttcaggtt aaattttaag agccctggct
40741 gaggaggaag tcctttaaat gagggggagg cttagaatct tacttttggt ttataccgct
40801 tatattcaac agtaataatt gattgtgatt attatgatca tgtttacttg gtccaagatg
40861 aatctgtgaa aggccttggg ttgatactaa cattcaagtg agaacatttg agaaaagtat
40921 acagaaaagg ctaaaacagg gtacttgaaa tgcaaagact tatcgtaaca gagaatgttt
40981 aattgttaat tggttttaac aagaagagca tcttaatata aatggatatt tttccttcta
41041 tttatttatg ttcctgttcg taagtgttga gaatgaagga agagctgcca cagacatttg
41101 gaagctggcg tggactccca gaagggctgt gttattctcc tatggacaga aataatgtca
41161 cagaatagta accacacaca gggtgactct gagattgtga gcaaatgaga cagagcaatt
41221 agacttcata atactgtcta agtatagaca gaaactgagt ctctgctcca cccacaaaat
41281 accaaacact ccctctcctg gctaaataga atgcctgttt tttctacacc aagtcaactc
41341 taacctcttc ctggtctccc tctctctaga tgagatttgc tgagataccc actggtagac
41401 ttgcccccca ctttctgcaa cacccaatcc agagtgggct ccatttcata ggccctcctg
41461 ccaatcaccc agccagaacc cagatattct aacaggttcc ttctcgcacc ccttatggag
41521 gtgccgatgg acctccacct gtgtccccct cactgagaag aatgaactca actggttcaa
41581 ccacctacag gagtgatctg gcagccccag cctggaaaca Ytgactctag agacttccat
41641 ggtttctgta gtgaggaggc aacaacaagc ccaggcagct gtgaggtggc cctgtcaggc
41701 ccctcaaagg taggggttgg ctccaagcag gcccctatct ctttcttctt tactgtggct
41761 tcggctgcat ctgtttcagc tggctcccag gcacatccat tccagccgct ggtctttgct
41821 cccagacttt atgttaacaa cgcagtggtt cccgggaccc cctgatgcgt tttcatcctc
41881 aaggtcatgt gtagaagcat cctcatcgaa agcctctcac ctccctcttc ctatttcctg
41941 cctctgtctt agggatcttg catccaggtg tctattgtct ggggccccca tgcaaccatg
42001 gtgacagtga ggccaataga ggcactaggg gagtgtgctg aggggtttct gggtgcttct
42061 tgctgtaagg gataagggtg atagaaggag ggctcccctt ccccttcttc ctgctttgtt
42121 cagtgtgtga ccccctaggc cagggcagct gtctcaacat catgagccat aaggaccagg
42181 agccagcaca gagcagagga gcaggagggc agagagagct ggggcctcag accacacctg
42241 acacctctga tgagctacac ccagggtctc gtcctcctta ggcgttgttg gcctggctcg
42301 gaatcttgtc tggggcatcc gctgagggtt gaagctggat actgcccctc gagctgaact
42361 aggacacagc atgtcctcYc tgggcaagcc tgcagtttgg tcaattcccg gaaagggact
42421 ttagttcctg gaggaggccc agacagggag ccctccctcc tgcctcagct ctgagccaag
42481 ctctccagat gcagaggaca gagactctca gcaaagcagc tccctgcaag tggctgtggg
42541 gcctcctcct taggcagaga gagagggaga gaggggtgaa gacaccgtgc tcctcctgag
42601 gaatctttat ttttattttc aatatttttt ttttttgctc tgttacccag gctggaacgc
42661 agtggcacca ccataactca ctgtagtgtt gacgttctag gatcaatcga tcctcctgcc

42721 tcatccttct gagtagtggg aactacaagt gcatcactat gtcagtcaat ctttttattt
42781 tcatttttgc agagatgggg tYgtgctatg ctgcccaggc ttgtcctgaa ctcctgggtt
42841 cgagggatcc tcctgcatta gcctYcgtgc attagcacgg gaattacagc cctgagccac
42901 aatacctagg tctcctgagg aatctttaat gacctgccag ggtcaagttc aacccсttaa
42961 ctgggcacac gggctcctac tttcctcact gctccctgcc cacacagtca ccttgggtcc
43021 aggtacacac tgacgccctg cttccctgcc tcccagagcc catctactcc ccccactccc
43081 attcccagcc tcaggccttt gctcaggctg ctccctcccc ttggaatgcc tcccccacct
43141 tccccaccat gtgcaagtca cccatcctca gaggctcagc caaagctgcc tggaccctgc
43201 agcgtccccc agccctcccc aggctccctg agcactctgc tttccggggg cagctctacc
43261 tctgtgtggc tctttgggcc tctgtctccc cgttaaactc tgagcatccg gcagggaccc
43321 tgcggtgttc ccagctgcat cccctatgac tgacctgggt ggcctctgcc ctccgtggca
43381 cagcctggct gaaatccgag gtgtgggagg cagcacttgg actaaaagca accagcctct
43441 ggacaacatg gcccagcttc agtttcactt tcagttgagg gattgaatgt ttttccttgc
43501 tatctagatg ctcatctctt tttggggaag aacaaataag accatgactc tgatttcatg
43561 tacttttggt ggctctatca ggtttctgag tttccattta ttttctcccg tgtcctcccc
43621 agtcttatta cttttgtacg aggRgaaaga aaaaagaaat ttctaaaaac acgcgagcca
43681 gctaacctct gaatcatagc catggtttcc aagcctcccc tcccccacgg caatggccca
43741 ggcctgcatc tgggcccagc tccctccggg tccatgacac tcttatgcag caaccaacta
43801 ctgccaggga cttttacggg aaagaggtca tttccttcct gcacttgttg tgccttagaa
43861 tacacacaaa aaagtcttgt tcttctttgt ctgtagtagg tacccatgaa ataggtccta
43921 catttaaaag aagaagaaac aaattcttct ccaaaaataa gtaaatcaat cccaaaggga
43981 tctaagtgta ccccagaaag aaaaaattaa aactatacta gaagaaatgt taaaaacaat
44041 gttgagttgg gaagaccttt cgaatcaaaa ctcaattccc agaagactta aagggaaaat
44101 tgaatagagt tttacctctt acaaattaaa aattagaaag aacatcattc aagagccaga
44161 ataaaatgaa agaataggaa gcacggtttg caatacttaa agcaaaggta actgtccaca
44221 atccacaaga aatgctcaaa tcaacatgaa aagactgaaa tcaagtttct aaaaatatag
44281 ccaggtgcgg tggctcatac ctgtaatccc agcactttgg caggccaagg cgggtggatc
44341 acctgaggtc aggagttcga gaccagcctg accaatatgg tgaaactctg actctactaa
44401 aaatatcaaa atttgccagg tgtggaggca cgcgcctgta gtcctagcta cttgggaggc
44461 tgagacagga gaattgcttg aaccccggag gtggaggttg cagtgagctg agattgcgcc
44521 actgcacttc agcctgggca acaaagccag actccatctc ataaataaat aaataaataa
44581 ataaaattta aaaatgggta aaagatagga tctgaagaaa tagaaattaa cgttaaccat
44641 atgaaaaggc actcaagctc acttttgaac catatatgta agtgatcaca gagacatacg
44701 tttttgaaac attagataag ccaaacctaa agtgactaat aatgtccact gttagggagg
44761 taaaggtggc acaagcccct tcactgttga tggagtgtga aaggttctgg tgtcctggag
44821 gatcgtcagc gatactcagc acactttaca aagtgccccc acgtcatggc acagcaccca
44881 ctcctaggtc tcagcctcca gggagcagag ggcaagtgtg cagggacagg ggctttagag
44941 ggtgccatgc agcgttgttt ctatcaccaa aaagggagaa atcactaaat ctcctgccat
45001 agggagtagg cgggtcccta acactatgcc cattaagact aaggaatgcg Yctggcacgg
45061 tggctcatgc ctgtaatccc agcactttgg gtggcggagg tgggtggatc acaaggtcag
45121 gagatcgaga acatcctggc taacatggtg aaaccccgtc tctactaaaa atactaaaaa
45181 ttagccgggc gtggtggcac gcctctgtag tcccagctac tcaggagggt gagctggaga
45241 atcacttgaa cccaggaggc ggagggtgca gtgagccgat ggtgctactt gcactccagc
45301 ctgggcgaca gagcaagact ccgtcagaaa aaaaagacta atgaatgcga tgcctcccтt
45361 aaagaaatga ggaggaaaag aagctgcaca gaatcaaaga cccttcgagg ccacgacaca
45421 gaggtgagaa acacagggct ccatgaagtg aaatctgaat agcaggtgtg agtgtgcgtg
45481 tgtgtgtgtg tgtgtgtgtg tgaatttgtg agtgtgtgtg tggcatatgc aaggtttgcc
45541 ttttatttgg ttgccagcat gtcaaatgta attgcacctg agtttatagc aaatgttccg
45601 tcatctgagc tctgaggagc caggcccacc ctctcacctg ctgggcatca cctggatctg
45661 agaagcccat gcaggcctcc agctctcccc actttcctgc agcccctgca ggcatctggg
45721 ctgcaatgtc aactttcccc tgactcgcac tgccccgcct cccacctcac cccagatggg
45781 gtcagcagcc tgctgggggt cagcagggta gggcatccta accccctMWg gcaggtgaaa
45841 aaccctgagg ctcagagagg ggaagtgact tccccagtgt cacacacctc ttggagggca
45901 gggggaaggt caacacctga cctcctgact ctgacattga agcaagaaag tataatgtta
45961 ataatccaaa atcttccagg tttgtttctg cttctacaaa cacagagggc taagctcacg
46021 ctcccacagt tatgctcacg cttccacagt ttgactttta atctcatgat tgtcaagctg
46081 ggatcatgtg gcagacacca ttctgcaaca tgccgttctc actgcccctg tctgggccat
46141 ctttcccact agggaacagc acagcttctt tcttctctat cacccagagg cctgaaaagg
46201 aaccaggtga gtgtcaggga catggagtcc tgtgatctac ctgctcaact tcctgaYatg
46261 ggaccctcag cgagtcactc atctcccatg gcctcagttt acccacaggg cctggctcaa
46321 gtaggtgctc agagtcaggg tgggtgaaac aacctctggg ttccttgtta tctcttttcg
46381 ctctctctct cttttttttt tttgagacag agtctcactc tgtcaggagt gcactcaggc
46441 tggagtgact ggtgcgatct cgactcactg caaccttcat cacctgggtt caagtgattc
46501 tcatgcctca gcctcccaag tagctgggat tactggcttg tgccaccacg cccagctaat

46561 ttttgtatt tttagtagag gcagggtttc tccatattgg ccaggcaggt ctcgaactcc
46621 taacctcaag tgatctaccc gtctcggcct ctcaaagtgc tgggattaca ggcatgagcc
46681 accgcgccca gccaggttcc ttgtatctca aaggaaatcc tctgtcccca tcgggcctca
46741 gcagcagcac cctggggcca gcctcagggt ggccaggcca caagaaggac aggccaggct
46801 ctggatcacc tcctgtcctc caccttctct ggttcctcca aatattctaa aggaatggac
46861 agtcccaaca tctgccttcc ttctggggtc accgggccat gtctcacgca ggctggaact
46921 agagctgagc cctaggaagc tttgtgaacc cagctgagct gattctgcct ctgtctctac
46981 tgtttggggg cccagagaga gggagggaga agtccagcta aaacctgggt gcaaattcag
47041 gctctgccat ggaccgagat gtgacctcag tcagttacct aacctctctg agctcatcta
47101 caaaaggtga atgacccttc tcttatagag ctgctgtgag gatgggggaga tggggaaggt
47161 cagaccaccc ctaacttacc aaggaagctt ctcttgaaag agtgtgagca agatccagct
47221 gttctgagct gtgtggatcc cacctccagc cgtgcatctg cataataacc agacacgttc
47281 tccagKctct gagatatacc ctggaaccca aaagggaaag tgaaagtcac aacttggcag
47341 cagctcctga tcaaacatgc aaaacaggat gcttcccagc cccacccttg gcccagtccc
47401 atccttggtc cagcagcacc cttggtYcca ccctcctgct gatccaagag cagccccaga
47461 cgtccttctt ggccaaccca cctgcctccc atacttaggg ttgggggttt gctgtgtctt
47521 caggaagaaa gacctaccac tgaaagacta tgagaccctc cagattactc cccaccccca
47581 ataacaccac actccccaga tgcccacctc actccctcca tgccatcagc attaccacag
47641 tccatcacac atgccgggtg gacgcttatt gcatattgat tgtattgctg ttggggttat
47701 tttgttgttt tgttttgttt tgtttttgag atggggtctg gctctgtcac ccaggctgga
47761 gtgcggtggc acaataatag ctcactgtag cctcaaattc cagggttcaa atgaacctcc
47821 tgagtagctg ggactacaag catgagccac tcggcctggc tacttttaaa aaatgatttt
47881 gaaggggtga aatcttacta tttttaccag gttggtttca aactctggcc tcaaatgatc
47941 ctcctgcctc agcatgagcc acYgttccag acctgctctt gtcgaacgcc atacctgtgt
48001 ggcttctgga caYttcctta tgtgggttct tctgccagga gcacctgcta ccttcacctg
48061 gaaaagtcac gctcctctat cagaatgcaa atctagcatt tcctcccctg agaagccccc
48121 tagcatcctc ttcagcagag cccactgccc catgttccac tcctgtggca cttttgttag
48181 gctgttgtaa ttggccacag tcccaggttg actgaacaaa gtggggggYg gatgtgggag
48241 taaaagacaa agacaaaaga gtatattagg aagaaggggt cNgggggttc cttgcttcta
48301 gtgaacaagg gctctgagct tttcaagctc tctgagttta ttaggtaaaa gagatgacga
48361 gaaaaaggtg gggggNtgat tgtcaggtaa ttgtcagtca gccRtttggt tcacagcagg
48421 cttgcaagac tgcatccttc aaataatagg tgctagatat cccaacagat aacttcaagg
48481 agcccagcac cagggagtga tggccctcag gaaaccttct ggcggcgggt gcagtgtgag
48541 tttacccaca tcccgcattc atgataaaca gtttgctgtt tgatcatata gcctccagtg
48601 aaatgctgag ttggtcacaa tccctttggc cttttcagct cccaacaggc agttaagtag
48661 acattagcag ctgggaggag aaaagagagc agaaagcctg tcactgcaac agctcctggc
48721 ccacctaagt tcagccccag agccattcta agaccccccct aactgatgaa gttggtggta
48781 tagtttgtgg ccagcacatc ctgaagaaag agaaactagg gcacaggtgg aaattcccta
48841 aaggggcaca tgcccagtaa catgaaactg tgccctcagg tcaccccaag ttcattatac
48901 catcattata ataaaatgtg gttttattat acaaaagccc cagagtgggc ttttctgtga
48961 aaattatggg taaaaacatg cattgcctaa ttttagttat ataaccataa actgccaatc
49021 agatgacatc gtcctttact cagacacagc cccaacctca acttctctcc acaaaccgca
49081 cagaagcacc ttaaactcta tatagagggg ctgacttcac ttcacagaaa ttagcccact
49141 gtccctctga gagtgtctaa ctgtgcttca atacactttg ctttgagctt gcattttgga
49201 gttagcctgc aattgtttgc tcactctcag aagaactgag attactggtc cagaactcca
49261 gctctgttaa tctcctcagt taaaggatcc atcccagcgc agaattcctg gtaatatctt
49321 agtgctgtga cctggatgat ttacgttcag catggctttt aatatgtgct tgattttagc
49381 ttggggaact gaaacatatt ttgctaaaat gcttcctttc gctttgggtc aagatgtttg
49441 tgctatgttt cgctgcttgg ctaagagaag ttgcaatgag tgttgtattt gctgcattag
49501 agagggggta tatcccccaa catgccacca ataactttta gagaagggct tagagaaatc
49561 ttctctggat ttaggtttaa ttagctgttt aacgtgatta aacaacttca ttaggttgta
49621 gtggccattg aagcccagag cttgttaaca gcccagggaa gctttgagag gcaaaatggt
49681 ttcttatgct ctggccacag agccaggaat caccctcaac tccctttaga ctgggcattt
49741 tggctggatg caatttttaa aattactgat taagagagcc gattgggatt cagatcatgg
49801 gatagcttgg ataaatgcta agtttagagg tttacagccg ggaacgaagg gtaggcccat
49861 agtgggtcat aacctttgga gaactcccca aatcgtcctY gcccaccctt ttgggcactc
49921 accccacccg aggccccctc cctgccacac tgccataggc agcagcacag cttctgaaga
49981 tcgttttgca atttttttc ctttttcttt tgaacagagc ttcatttctt ttctttcttt
50041 ctaaccactt ttattcagaa tctaggggac aaatgttttc agcaaggctg tagtacccac
50101 tgtaggaagg ggtctccatg ctgggcacag ccatggtcca cacataggta caacttgaga
50161 gacacacaga ctgccccagg ggaggcggag cctccatcta aagagcaggg gacactctct
50221 cctggtaatg gtcaaggata ttgggatgtc gagactacct tgatctcagt gctgtgtatt
50281 ctggattgcc agctgcatcc acactcctat gtaaaaccac agagggctgc ccaaactgat
50341 gatgtctcca cagacaaata cacttgtgag actttagtgg acctcaagcc tctctttaaa

50401 agactgccaa aaaataagta atgcccactg actttcccta gtgccttttt ttaaatatta
50461 gctttaagtt ggattgttat agttattctc atatatttgc agttcaattg cctcctcctg
50521 aatgggtgaa ataatttttt atgctaaaat agctgtttta aatgcagtaa ccaatatttt
50581 tgctttttaa actagattaa cccttgcaat catttctcct aatgtttttt tccaacaccc
50641 atatttagct accatcaaag aagttttaaa accttgttta aacacaacat caacggactg
50701 cacaacatca cagaaactgt tgaatctgtg tttacacctt ccctagcaga ggtctatcag
50761 ccatttccat tcctgcttct gagtggaatc aatcagaaac atttccagcc cccacttctg
50821 ttaaccgacc agaataagct gtcaatccta gggtcgagcc cacagtcaat tgcatactcc
50881 ttaaaagtct ataccaacaa acctgttaca acctccgagg acctgtcgtc tatctgtaat
50941 actatctgac taagcactcc tttgtcttct atcttcattg aggatgaggc cattacttgg
51001 atttttcct aatatttcca aggcttgcaa tcacacttgt accctatttt ccacccaatt
51061 atgtgtcagc ttagtcactg acaacaaaaa atgcaaacat acctgatatt ctaattcttc
51121 tttgtgtttt attcatttcc caagcccccc aacagaacca ctatgtggaa agaagccagc
51181 tattctggat ctaccctttt ccccagaccc acaccagaat atcattttaa acataacaac
51241 ttctggtttc agtctcacaa cttcgggcaa aatcattgca aaaaactgca gctccccata
51301 atttcacaat attggttggg atcccatcac tccttcctct aattgggctt ttgattcctc
51361 tttcattctg tcaagtactg cctgtgttcc cccattacat gaactagcct tgacttccac
51421 ttctgagggt attaaagcag acttacacag catggaaaaa ccctacttta gtatcactgt
51481 tgatatttgt cttaacatga caaatcacat ttttgtatag aaacaaggca tatctctgct
51541 tgtctgctaa ttaaacgggg atttggccct agtttacatc atccccaaac tcactcttat
51601 cgatagcatt cccctcccta tacacgctca atctggtaaa actgaaagag gagttttaac
51661 agcactgact ataattggcg ttctggctgg tattggcacg ggaactggag gacctcctta
51721 actccatttt tactgcccag aaactcttct taaaaattac tcaacaaatt attaatctaa
51781 ctaaccaggt caatacccтт tcagaacaaa ttgcttcatt agcaagcata gtcttatgaa
51841 attgatgagc ttttaatttg ctcacagcca atcaaggggg agcttgtgca acacttgata
51901 aacattgctg ttttcatcaa ccaatcaggt aaagtgcaaa ccaacttaaa aaaattacag
51961 acagggccaa caagttaatg gaaaacagag tttcaggagg ctttgactgg tggaaattta
52021 acaactggtc ctggttttcc tggttagccc cctttcttgg tcctaaaaca gctttataca
52081 acttatagct ttcaaacgat gctcatttga tttacaaggt tgttttgttt cttcccacct
52141 agaatccata aaactccaaa tgattctaca caaggttacc gcgaggtctc aaccatgtct
52201 gatatctacg tgggcccctt ggatggccac cacttttctc tgcccaacct accagctaca
52261 agagagcaaa aaagagaacc caattattac cactgcctct tgtcagcagg aagtagccag
52321 attaattgtc accgatctac ccaaaaattt ggcttttttt ctctaaaggg ggaaatatta
52381 ggtagttaaa tagacattag cagctaggag tgggtaagaa aagagagtag aaaggctgac
52441 agtgaaacag cccctggccc acctaagttt agccccagaa ccgtcctaat gccatcctaa
52501 tggatggagt ttttggtaaa gtctgtggcc agcacatccc aaagaaagag aaagtaaggc
52561 acaggtggaa gtctcctaaa ggggcacatg cccagtaaca caaaaacatg ccctcaggtc
52621 accccaagtt cattatacca tcattataat aaaatttaca tgtggttgta ccctcctttc
52681 cccagagtgg gcttttctKt ataaatgatg ggtaaaacca tgccagttta attttagtta
52741 tataaccata aactgccaat caaatgacat catcctgtta ctcagacaca acccaaacct
52801 tgactcctcc ccacaaaccc cataaaagca ccttaaaccc tgtaaagagg ggcttatttc
52861 acttcgcaga aatcattccg ctctccctct gagagtRtat tactgtgctt caatacactt
52921 tgccttgagc ttgcattttg gtgttagtct gcaattcttc gctcactatc agaggaactg
52981 agattgctgg tccagagctc cagctctgtt aatccccacc tctgttaaga atccatccca
53041 gtgcaggatt attggtaaca cttagaggga aggggcctaa gcccactgct cagccacagt
53101 cccgtggcta agaaggcgtg ctgtcggatt gctggcggac ttttctgctt taccccaggc
53161 agaagggttt tgagaggggc tgaaccctga ccaacatgac cctttcactg ggaactgttt
53221 catcctgggc aaatgtccct gcagcctcgt ggccaccacc ccatacacct gccagatttt
53281 cagggaggaa caggagggtg attacacaag aactgtacga tacactttga aatataccca
53341 cttctctatt taattggaat ttgataacca tctcttaaat gggccaggag tatgatctga
53401 gtttacaggt gagRaaactg aggcaaggca tgaggctgtg gctttcaacc agcaactcgg
53461 ggaggagcca agatttgtaa accaagWtgc gtctgagaca gatctcaatc cgtttagtgg
53521 tttattttgt catggttgag gatgcccaca ggaaaaaaaa aagagacaca agtcacagtc
53581 ggatctgtgg cccacaagtt ttcaaaagag ggttttaagg gcttcaatat ttcaaaggta
53641 aagagtggga aggaggagga gggatgaaaa aaaaagacag gacaggtagt gagaagagtt
53701 gtcacattct tgtgaggctt tgattagcac ttcctgaatc tatgtgctgc atgtaaaagg
53761 catggataga agaacagtaa ttatgcattc atctcatgct cagtaaatgt gtattttaca
53821 gaagataaaa ataaacaggc agggcttggt ggctcatgcc tgtaatccca gcactttggg
53881 aggcctaggc acttggatca cctggagtca ggaggtcgtg accaacctgg ccaacatggc
53941 aaaacctcgg gtctactaaa aatacaaaaa gtagctgggc atggtggtgc acgcctgtag
54001 tttcagctac tcaggaggct gaggcaggag aatcgcttga atccaggagg tggaggttgc
54061 agtgagccgg gatcatgcca ctgcaatcca gcctgggcta gagagggaga cKccatctca
54121 aaacaaaaca aaacaacaac aataaaatga acatagatta gaggaagcag tcacatatgc
54181 atttgtgttg aagtgagcag cagggtaatt tctagtcttg gttttgtcat gtacctgtca

54241 aggtaagtta ctaatttgca ttgtcatggt aaaattccac aggctctgtt ttaagggaaa
54301 agatgttggg gctcacaagg aattttcttg tgggaaatcc atgaggaaga ccacctgggg
54361 aatatgttgc cttctgtctt ggcagatacc tgtttaggga caaaaagaag ggcagttttt
54421 tgactgattc aatttccctg cttaacggtt ttctttggca tagagagttt ggggccctga
54481 tattttattt tcctttgaca ttatctctcc cttttacaaa atctttttga gaaagcattg
54541 gagatgcaaa tgagtctctg gtcatgagtt tcatctgatt cctgtcagat gaagttgtct
54601 agtttcagtc tgtagggtgt tcagggaagc acggttttaa tttctggtga ttccgagtga
54661 gaaaaatgag agaaagaaga caagaatgat gcttcttttg agacttgcag ccaagaaaga
54721 attcacgagc cagcctagat gaattttgga caaataataa aactggaata caatggacaa
54781 agctagaatc taataacagg tattctacag tttcttttaa aacatatttt tttttctctt
54841 cagtcctcta attctaccaa agacaaatca tggtaggact aatttatttg aaaaataagt
54901 tgtagtctta tgatacttgg cctgattact ttcataaagt gcagtaagaa caatgattta
54961 ccatatgggt tcttcttaag aaattggatt tgctggaaca tcttcttttt cataaggaat
55021 ctcagattac acctttttaaa gccttgagtt cagccaagta tttatctgtg cctgcaggtt
55081 tctgtatgaa ttgggtgagt tcctctcttc ttgagattcc aagataactt gggggttcct
55141 aggcctgtca gaaagtatta ttttccttac ttaccccaca tcccacatca ggaaccctgc
55201 atagttaaag aatgaggcca attttttgat ggggctttca ttggctctat aagtcaacct
55261 ctattcctca aagctgcctg ctcatatgtg aaaatatgcc atgccagtca aagccttgat
55321 aaaataacta gtgtctccaa tggtgtccta ttgcaaagga aaacagattg ttactgaact
55381 tatgcaaata attctattgc cataaattaa gaaaactcac aaatagtttc caaattctgg
55441 agaaatcaaa tagagagaaa gaaatgtgct ttaacttttg ctcacaagag tRtatttcat
55501 tcaactgtca aaagctgtac atagctcaca agatatcaac gtttccttga ctctgaagaa
55561 gaaaaaaaaa agaatcaggc atgtttcaaa cacaaagtca tctaaaaaat tatttcagtc
55621 ctctattaga tcagtcctaa gatagaaaaa aaatcaaaaa tgttattttg gagacttgcc
55681 gccaggaaag atttcaggat ccattctaga taaattttgg acaaataaca aaactggaaa
55741 aacaatgcac aaggccctaa tctagtaaca ggtgttctat aactcgttcc acctgatatt
55801 gggttagcaa tcctcatgaa cataacaact ttttaattag agtcctggat tttttttctc
55861 tagtccaatg gcatagcctg tagctgacca taaactggta ttgagaagaa ttaaagtaaa
55921 acagtaactg tggatagcaa aagtattaga gcagccttgg caaacacaaa attgacaagg
55981 aaatttggtt atttctgtga cacacaacaa cttaccgtaa aaatgattat tattactaat
56041 aacatatatt gctgtgtgcg gtggctcaaa cttgtaatcc cagcactttg cgaggctgag
56101 gtgggcggat cacttgaggc caacagtttg aaaccagctg accaacatgg caaaatccca
56161 tctctactaa aaatacaaaa agcagcttgg tgtggtggtg cacacctgta gtcccagcta
56221 ctcgggaggc tgaggcatga gaattgcttg aacctaaaag gtggcggttg cagtgagcca
56281 agattgtgcc attgcactcc agcctgggtg actgagtgag actctgtctc aaaacaaaca
56341 aacaaacaaa acaaaaacac catagaggaa gctataacag aattataggc atctcataca
56401 attttggaac acatttattt atatataact caaagaaagt tacaacacca ttttacatct
56461 gtcctgctag acccgtgcta gacacttgca agctcggaga gcacagtctt ggagagagta
56521 aagaaaagac ttggagagag cagatgagac ataagcttta ttcagggaac ttacatacaa
56581 ggagtccagg agcagcaggc tggacagaaa gcccctcaca tttgtaaaaa gcatgcagtt
56641 tatatagtat ccttccccaca gcaacctcca cttagccacc tccacccggc aacctccact
56701 tagcccaaaa caaagggcct tggttccctc caggacctgc ttcccaagga ctgggccagg
56761 ttacaggtgt tcttcatagg tatagtgtaa acatctggat tggccattcc tggggtccat
56821 agtttagaac aatgaacaaa cattcaccaa ggaacatagg gccattctca ggggatgctt
56881 gaattattgc tgtctgacag gacatgccca ctaggcactc cactccagta tagccttcag
56941 tggccctctc gtaccaagta agatgaatat gtctcttcaa gacttcaggg gacctaatat
57001 ctaaaaaatt aatgaggtca aaacactgga tttgaaattt gattttagaa agtttgtcaa
57061 ataccaaagg ttgaaaacac ttgatatcac aaaataggat tacgctcact ataaagtaag
57121 tcattcgtgt ggccaaagag ataacttaaa tatttcaaaa acatagaaat gtttacacct
57181 tgagagagag aagactcagt ttcccacaca ataacacctc ataaagacag catgggccgg
57241 gcatggcagc tcatgactgt aatcccagca ctatgggagg ccaaggcaag cggatcacct
57301 gaggtcagga gttcgagact ggcctggcca acatggcgaa accccgtctc tacttaaaat
57361 acaaaaatta gccagacgcg gtggtgggca cctgtaatcc cagctattcg ggaggctgag
57421 gcagggagaa ttgctggaac ccgggaggca gaggtcgcag tgagccgaga tcacgccatt
57481 gcgctccagc ctgagcaaca gagtgagact ccaccgcaaa aaaataaata aacagcatga
57541 ggccagtgga atctgtttct ctccctacca gctttttttgt agtttattca aaaagcaaac
57601 agacgtcttt tattatcttg taatattaca tgaaaatctt gttcaaagca gcaagcccaa
57661 atttaccttt gcttaacact aaagttaatt ttaatatctt agaaacaaat ctgtcaaatt
57721 ttaattagtt acatcttaag gtaaaatatt ttataaatat tttataacct tttcaatttc
57781 tattaaaaag cagatcaatg ctccaagaaa gccctgttat tctcatacag tggctcagat
57841 tctgaccttt catcagtgtg gtatttatat taattttcaa tttataggaa aagtaaataa
57901 tccccattac aattttgcca ccttgatcac aaacaaattc tttttacaag aataatcttc
57961 cacacacctt ctataattgg ctgaaaactt tggtttggtc ctgtttttca tttaacttga
58021 aataatcctt aaaaacttct aaactagata acattattat tattattatt attatttttt

58081 tttgagacag agtctcgctc tgtcgcccag gctggagtgc agtggcacga tctcggctca
58141 ctgcaacctc cgcttcccag gttcaagtga ttctcctgcc ttagcctcct gagtagctgg
58201 gactacaggg gcgtgccacc acacccggct aatttttttgt atttttagta gagacgtggt
58261 ttcaccgtgt ctactaaaga tggtctccat ctcctgacct cgtgatccac ccaccttggc
58321 ctcccaaagt tcttggatga caggcgtgag ctagcacgcc cggctaatta cgtttaacaa
58381 aacccacatt cctgtgcctc ttataacgtt tttaccaaaa acacatccta ccttttgtca
58441 caggtataac tgactgggac caatattgtg gacagtaaaa taatttaccc aagacagtca
58501 tgggtaaaga aagacagatt tattagagaa agtacaaaga cacattgcaa aagtgcaatg
58561 gtcagcacag caaagaaggg acgtctgcaa agtgtcaggg gctgaaggga agtttcatag
58621 ggttgtactg gaggctacct gtggaatgag gtgcagctgg ggctacatgc ggagtgaggt
58681 atttgggaac aggatgtcat gctagttggt tgtctgtgat cagctgtctc tcagaacaat
58741 tgttctcccc cacctggggc cccttcctcg ttgttgctta cttatcttat taggactcca
58801 cacaaattac tccttatcag attttagcca ggacaaacag ctgatatttc tggcttttta
58861 acttttatac aaaagataac ctcctgagtg aaaccaataa cgcttaacta aggtcattac
58921 tgaaccatgg atgcatggat tccccctgtt ctcaggctac acactagaca gaaagtgagc
58981 taagctattt tcactgtatt cttatctttg caggtgtaaa ctcaaccttt tccccacggc
59041 acttccttgg gttattctga aagcctctgc agtactctga ttatcccgat gtatatacta
59101 catgaaacac tgtctggttt attagttcat ttatctcact aacagcaatc atactaatac
59161 gtttttatga cctgagaagc ctttgcttca aaagaagaag ttttagtagt tgaagtatca
59221 actacaaatc tcaaatgact ctacggttgc ccaccaccat accacacatt tgaagaacca
59281 gcttacgtaa aatccttgac gagaaaggaa ggaatcaaac ctcctaaaac tggttccaaa
59341 cgagcttcat aatcactctg actttctcaa ttaaaatgta ttagtaaaaa taattacata
59401 actttgtcaa agttaacttg taggttaaaa ttatttatct ctttatggca tatccattta
59461 aactaggctt ccaagatgct atgtcaccta tcatagaagg aagaactcct tcattttcat
59521 gaccacaccc atgtggtcat tttcttttct ctgttttttt tttttttttt ttttttttga
59581 gacagagttt cactcttgac tcttgttgcc caggttggag agcaatggca cgatcttggc
59641 tcaacacaac ctccgcctac cggcttcaag cgattctcct acttcagcct cccacgcagc
59701 tgggattaca ggcatgtgcc accacgcccg gctgattttg tatttttatt agagacagga
59761 tttttYcatg ttgatcaggc tggtcttgaa ctcccgacct taggtgatcc gcccgcctcg
59821 gcctcccaaa gtgctgggat tacaggcgtg agccaccacc gtgcctggcc gtgatcattt
59881 ccttaattcg ttcattagta ctttatatta cttccctgat actaacaaat taacttgcac
59941 aagactatag atgcccaaga agtggagaca atatgaacta tgttgcttgc tagtatctta
60001 attttaattg ttctcccatc attatgtgtt ctaatacata atagatgaaa ttaataatct
60061 ctccctttct gtcaaaacta tagggcatca atgatattga agctacaaat atacagatta
60121 tgaagaccta agttttgact cctacataat tcccacaaca gacttaaaat caagagagct
60181 gtggttactt gaagttgaca gttgagtagt tctccctata gaaatattaa ttcatatgtt
60241 aatttcatca gaagatgtcc tatatttgtg agccatacca tcacatcaat aggcctaaaa
60301 atcgttRcaa tcccagggca tttaaaccaa gcaaccttaa catctatacg accagacctt
60361 tactacagac aatgcccaga aatctgcaga tccaaccatg gctttatacc cattgttctt
60421 gaactagtcc ccctaaagca cttcgaaaat tggtcagcct ctgtatcata atatcactgt
60481 gaagctaatt agcattaacc ttttatgtta aagaccgaga gccaggattt ctccacatta
60541 aaatgtccca actaaacaca tctacattat ttactatggt tttatgtata atcatatcac
60601 tctttatctt atttcaatta aaattctcaa aatttatctc cccaataagc tcaacactaa
60661 aattctttga atcacagaaa cttgaaattc cttgagaaat Raaatgaatg gaaatgtatt
60721 tgcctttcat tgccctaaca gtaatacgaa gacccgtagt tataatctcc tatttccaag
60781 tattttattt ctatcatcta gtcacctgat taacaactgg ctaatctcca ttcaacagcg
60841 actaattcaa ctcatattaa aacaaatgat aataatacat aataccaatg gatgaagctg
60901 atcccttata ctaatttcat gtattctgtt tattggctca actaatttac tRggaggctt
60961 cccactcaca ccaactactc aactctcaat aaatttaggt atagctattc cagtatgagc
61021 aggagcagta attactggtt tttgtcacaa aacaaaagca tcccttgctt atttcctacc
61081 acaaggaatg cctatttaac ttctctcttt tctagtcatc gtcaaaacca ttagcctttt
61141 cattcaacca atagccttac ccatatgatt gacagccaat attacaactg taaatgtaaa
61201 ttatacattt aactggaggt gccatcttag tatcaacatc tattagccta cccacagcct
61261 cagtcccatt tattaccttg attttactag ccgttcttga atttgccatg gttttaatcc
61321 aagcatatgt ttttacaatt tttttttgag acagagtcta gctctatcac tcatgctgaa
61381 atgtagtgac atgatcttgg Yttactacag cctccaactc ctgggctcaa gcagtcctcc
61441 agcctagcct ctggagtagc tgggactaca gtttcatgtc accatgtctg actatttttt
61501 tattttttat tttttggtta gatggggttt tatcaatgtg cccaggctgg tctcaaactc
61561 ttaggttcaa gcaatctacc tgccttggcc tcccaaagtg ggattatagg tgtgagacac
61621 catactcagt ggtttttaca cttctagtga gcctctactc acatgatagt actttacgac
61681 ctacaaaacc caagcatata atagagtaag cttcagccct tggccattca caggggcgct
61741 cataactact gtaacacttg ttgaaaggca ctcattacaa catctgtctt agccatatga
61801 ttccactcga actcaattac tctattaacc ctaggcctag taactaatac attaacaatg
61861 tatcaatcat aatgagacaa tatcagagaa ggcacattcc atggccacca cttgtcaatt

61921 gtccgaaagg gccttagata tggaataatt ttatttatcc catcagaagt agttttctct
61981 gctggattct tctgagcatt caagcctggt cccaatccca gaactcggag gctgctgacc
62041 tccaacaggt atttacctcc ttaacccctt cgaagtgccc ctactcaaca tatccattct
62101 ctcagctttg ggagtatcaa ttacctgagc ccatcatagt ttgaaagaag aaaaccaaaa
62161 acacatactc caaaccttat ccattacgat tgccctggat atttacttta cactcttcca
62221 agcctcagaa tgcttccagg cacctttcac cacctctcat ggcatatatg gctcaacatt
62281 ctctatcgct acagacttcc atggacttca cgtcatcaat ggctcaacct tcctaaccat
62341 ctgccttgtg catcaacgaa aatttcactc atcagactYt tgccaaattc cattgcacat
62401 ctgctctaca atcaaagaca ccttggatct aattctcctt ctaatattac tactcatatc
62461 agtactgttc tcaactgacc tattaggaga ctcagacaat tacactccag caaatctcct
62521 caacacacca cccgacatta aaccagaatg gtgcctttca tttgcctatg caattctaca
62581 ttaaaccaga agggtgtctt ctatttgcct atgcaattcc acattaaacc acagtagtgt
62641 atttcatttg ccaatgcaat tccacatgaa accagaatgg caactttcat ttgcctatgt
62701 aattccacat taaaccagaa tggtactttt catttgccta tgcaattcta cactccattc
62761 ctaatgaact aggaggcatg ctggcccttg tcttctccat tcttatccta gctgttgttc
62821 caacacttca catatctaaa caacaaggca taatattctg accattaagY caatgtctRt
62881 tctgaattct agtagcagac cttcttacac ttacatgaat cgggggccaa cctgttgaac
62941 acctctttat tattatgagc caagcagcat ccattgtgta cttttccctt atccttcttt
63001 ttataccaat tactagccta atcgaaaacg agctactcaa atgaagaagt ccttgtagta
63061 tcattaagtg actctggtct gtaaaccaaa aatggaggac ccggctcctt gggaccattt
63121 ccaggaagaa gcttcttgct ccacaaccag cacccaaagc tgaaattcta cttaaactat
63181 tccctgattt ttgtacttga agtgcaatac atttatttta actgctatgt cagattggaa
63241 aatccacata aattactgtt gcaactatgt acatcatgca ttcctgcttg accccatgaa
63301 ttaacatgct agtactagat gtgcctgatc gtatattgta tatttatgta tcgtacctta
63361 aatcttcacc tccatgaata tttagcagga gtaataatgt cacagcaacc tgtcatacat
63421 tcatcaaaca ttctgcagaa agcgattaac atacatatat atcatccagt ataataatcc
63481 cttaatattg cataggacWt tcagttattt atgaatcata gcacattcaa gtcaaatcat
63541 ttctcagcaa cacgcttatc acctcctaaa aacattctta actaacaagc tttgagaaat
63601 cagcagcctg cttggaaagt tctatccttc ttactcccgg cccatgacac ttgggggtaa
63661 ttagactgca actatacctg gcatctggtt cttccttcag ggccatggag actaagatcg
63721 tccacttgtt cctcttaaat aagacacctc aatgatttag tgacaatcac ctcatggtca
63781 cttgcagaag tactgacatg tatttggtag tttttaaatt gggagatgct atgactcagc
63841 atggtgggag cctgaaccca gccaactcca ttgtagctgg agtttaatgg aatattacga
63901 gtcaacactt acaaccacaa ggtgctaatt tattcatgct tgacagatat catgaaaaga
63961 aaatgcacgt tcaagctcac cctttctcca aacatgtatc tgatttactc ctcaaactcc
64021 cctgccccct tttccctgga ctaacctaat tatgccaaaa aaaattgttt gttcttgcca
64081 agcccaaaaa caagatatgt ataacatcaa tgacagccag ggctcacact ccaaatcaaa
64141 ttgtaatctc aactacaatt acattttagc caaactttca attaaaccta tttcaaaaac
64201 acatgattgt aaaataaaaa ccaggctatc taggctaaaa atcccacttt ttatttctaa
64261 agttactaaa tattgcccta ataNctagca tagtacttca acccttcatt ttgtaacaga
64321 gagaaaattg tctagatccg aattaatgta gcttatgata ttaaaacaag gcactgaaaa
64381 tgcctaggtt agttcatgta actccatRaa cacaaaggca tagttctggt ctttttattg
64441 ggttttaata agattacaca tgtaagcatc tggattccag tgaaaacgcc ctccagatca
64501 tctaatttca aaaggagcag gtactaagtg cacagcccta catcataaca ccttgctcaa
64561 ccacgacccc acgggaaaca gcagtgataa aaaaaaaaaa aagcaataaa tgaaagtttg
64621 actaagctat actaaactat tagggttggt aaatttcatg ccagctaccg gagtcataca
64681 attaacccaa actcacaaaa acaagcattc agcgtgttta aggcaattct cataataaag
64741 ctaaacttta aaccgagccR tgaaaagctc aagctaaaat aaaaagaaac tacaacaatg
64801 attttagtat cctgagaaca tgatagctaa ggcccaaact gggactaaat ctctcactgt
64861 gcttagccat aaacttaaat aatttaataa acagaattat ttaccagagc actataagca
64921 atagcttaag cctcaaagga catggcgatg ctttacagct ctcgagaggg gcctgttcta
64981 tattcaataa gctccgatat acctcagcat ctcttggtaa ctcagcctgt atactgctta
65041 aacaatctca caggaagtca ccccatctat taattattct tttatcagtc cccttgataa
65101 gtgggtcatc cactttatca tctcttattc ttgcacaata atctgcatga ttgattttac
65161 catcatgacc cctgtccata atgtaattca tctcaacact agcagaaact aaccgagccc
65221 catttgaccR gcagaatgtg atttagagct agtctcaggc ttcagtatag ttatttattt
65281 aattttttttt ttgagacaaa gtctcactct gtcacccagg ctggagtgca atggtgcgat
65341 ctcggcttac tgcaacctct gcctctgggc tcaagcgatt ctcgagcctc agcctcctga
65401 gtagctggga ttacaggtgc acaccaccat cctagctgat ttttgtattt ttaatggaga
65461 tggggtttct ccatgttgac tgggctggct ctggaactcc tgacctcatg tgatccacct
65521 gcatcgacct cccaaactgc tgagattaca ggagtgagcc attgcgcccg gcccggcttc
65581 aatgttgaat atgctgcagg ttcattggtt gtttcttcat agcagaatat actaacatta
65641 ttataataaa tgtcctaact actattattt ttccaggagt actacatgat gtctttatag
65701 caaaaatcta tacaattaat tccattacca aactctccta ctgagaaccc tgtttctgtg

65761 aatcagggca tcagacctat gagtttgata tgatcaactt atgcatcttc ccgtgaaaaa
65821 actttctatt ccttactcta gccctatgca tacagaatgt ttaaatatct gtcctaatat
65881 ctagtattca gccacaaaca tgggaaatat gtctgataag agagttactt tgacagagta
65941 aattacagag gtttaaatct tcctatttct agaattacag gaattgaaac tagtcctaag
66001 aagtcaaaat tctccatgct atctgataca ccatgtccta tagtaatagc agctaataag
66061 taagggcagc ttaccctatg gtaatggcag ctaaataagt tatcaggccc atacccaaaa
66121 tatgctRgtt tataccсttc ctgtactaat cagttctcta atcttcctca ctatcctatt
66181 cactattttt atagaaactg taattatagt aatcgcctta cactgtctct taatctgaaa
66241 ggattagaaa taaatatgct agtcataatc ccataaaaag gccaaaccca ggccgggaat
66301 ggtggctcac acctataatc tcaacacttt gggtggctga ggtgggtgga ctgcttaagg
66361 ccaggaggtc gagaccagcc tcggcaacat tcatagggag tgaggccatg cttggtagtg
66421 agtgactcat tcaaagtttt catggtttca gggagattct tagtctagga cccccaagag
66481 ccatcctggg acagtgctca tatgtgggtc tttagcaggg cactcttcca ctgaggtctg
66541 ccccagtggg ttacatgagc tgtggaacct ccagtctatg gattgttctg atgcccagtg
66601 ctggaaatca tggcccatga aagggggccc tgaYctctat ctgtcctttt agccatgcca
66661 ttcgtgacac taagagcggt gagacRcttc atgStctcca gttgggtggc acgttttaca
66721 ctgaaggcct gtagtagcct cattgttgta tccacacagg ttaagtcata ccgtttccct
66781 ccgctctggg gcgggggggc caataaaatc gactcgccag tcccgcacag gttgtatggc
66841 tcgatgtatg tgacccggtg tggagggaac ctttctagtc gcagccatga acaggtcccg
66901 gggttctgaa cagctgctag attatctgca cagtgggcag ggaagcctac agtctttgct
66961 ttggcccacc tcatggctgc cccctggtgg tcgctcttat aaggtagcca catggcagcc
67021 tcccacaatg ggctccagca aaccatggga acatgcttaa gggcgcctgc ctcctgattt
67081 ccaggagaca gcttaggact atgtgcatca acRtggtaca ccataagatg cacatcattc
67141 ttttgtcacc tgacgtggat attctgccac atggacaggc cctcaaaggg gcctgcccca
67201 aacctcccaa ctgtaagcag cccattgact gctttgtcgt tacccccatt ctatccRtat
67261 ggtgtcagta tccatttgca ccgtcagcac tgaccgtaag tgttaggtgc tcttgctagt
67321 tccatgtgtt tgcctggcag atatcgtaag ggtcgcctca tatatggtag ttggcatggt
67381 aggtggatct attgccatgt ccRctctttc aagcacgtcc tgtagtactt gacttaaagg
67441 actagtggac agggcatccc tttgttgcag agaggcatgg catttctgta ggatgtgtgg
67501 ttgtgccacc ccgggaggcg ggcttggcta ggagaccttc cattcaccct tgtataaggt
67561 aggcagtctt tactgggaca gtctgctttc tggcgatggc ctctgctcRt tgcaatgtcc
67621 tatacatggc caatagctgc tgctccagca ctgtaaagtg agattctgct cccttccaca
67681 gttgagactg aaaccctaca ggtaccattc ctgtgggttg catttgcgac aaactccaac
67741 tcaggtcagt ggcgtctcta gtgacttcca gtacaagagg gtgctgtggt catgaggctc
67801 ctgttgcttg agcttgtttc acccatattt tgttttgtca gaagcctctt gctccatatc
67861 tgtccagtcc catttgtttc acctttcttt attagagtgt ataatgggca aagggattgt
67921 gccaaatgag gaatgaatat cctccagtag cccagtaaac ctaggaaaac ctgcagttgc
67981 tttactgtct gaggaacagc atggatgcgg gcttgatggt cccctggggc ccatcggagt
68041 ctcttcccag gaaattggag gacctgaagg aaaaccctta gaggcagcag gcagtggagc
68101 tgaggcctga agaggagctg gtttcccact gctgacaccc ctgaggcttt ccggatccct
68161 ggtcttccct aaacctggct ttaaaattgc ctgtgtgtgt gtgtgtgtgt gtgtgtgtgt
68221 gtgtgtgtct gtccatacat gtgtgtgtgc acacccatag atttctttat ttatttaatt
68281 ctgtgagctt tctctctgga ttcttccaat acatgtctct actgcagagt tagccagagg
68341 atttaaaaaa gctttcgctt cacttttgtt cttgcatttg cttttgactt tgttttgctt
68401 tgcgctctcc ttgcttcNaa aaaaaaatgt taactgatac aaacattgtc ccaaagagtt
68461 ggttgcagat cacaggccct cagatgattg gcatatttgg aataagtttt caggctgaga
68521 ggagggcaga gcaacagtga gccaccccca accccagcat tccggggagc acctgctgtc
68581 atcttgtgga aaggggcaag gaacgctggS acctgaagca ggataggtca cttcagacaa
68641 atgcctggct ttgacatttt tcctgaaggc gtctgataaa ttgtggctgt tgatattttt
68701 ggctgacttt cctatgaaaa gcaaattctg taggtttctg gttacatgat aaatttcgtc
68761 ctcgccagat cttctaagcg cacatgaggg ccctgatagg gagagagtgg tgtgttagct
68821 cctcccctgc tgtaacaaag caccacagac acagtggctc aaaccaacac gaagttcttc
68881 tggagctcag aagtcctaaa atcaaggtgt gggcagagct gcgttccttc tggaggctct
68941 ggaagaagaa gccattccat tgcctcttcc agctcttgga ggaaactgct ggccttggct
69001 cctgtctcct tccttcctct tcaaagtcag ccttgtagca tcctccgatc tctcccagtc
69061 tctctcccct tctatcatct catctccttt tctccctctc accctactgt ctctgcctta
69121 gaaggaccct tgtggttaca ttggacccac ttaaagaatc ccaaataatc tcccatctcc
69181 agatccttat ctggacgata actgcaaggt cccttttatg atggaaggtg atataatcgc
69241 aagtcccagg gttaggacat gcacatctgt agggggctgt tattcagcct aaaacaagtt
69301 gcatcaagaa aattacgatg gggatatcaa ggagtttgaa gaatcccaac tcgtagccat
69361 gtctaagggc ttttctgagg tacaggcccc gctagaggaa gctgctccaa ctctgttcct
69421 gtctgctcct gccccactct ctgtttctag acagcctgcg ttcaagtcat tgtggcccaa
69481 gactgaactg cagaatggaa aggagtgagg cgctgcagca gaaggagctc gctgtcaaag
69541 ctccaaagag ccggagacca tttgcagggt ggggtctgag ggaggttggc aaggaaaggg

69601 gagcatcatg tggattcctt tccatagtYg tcatgagcta acacacttcg ccccaacagc
69661 tggcttagat tctaattgtg cccaggactg gccatgcaac ccttgagtaa atgctggcca
69721 cctctctatg acaccaaaag gccagaaatt acctggcatg ataaagaaga aggagtcaag
69781 tgtataggag accggagggt tagactggat ttctcaggcg ctttctgccc tgactgtgtg
69841 ctgccagggc ccaggagatt cccctttcct ttgatgagaa gaatcagtgg atattctcta
69901 gagctctcta gaatggggga ggggctagga ccctggtctc tttagggtgg gagaaattcc
69961 aagaggcaaa ccctagggtg aggcaggtgg tggccagaag aaaagagtgt ggacaccagg
70021 tcaRgaagca gggagggcgc atgatctggg ctcagggacg ttccacatgc agaatccagc
70081 aacagagctg gtggacagtc tactgacttc gcaaaacccc cagtctggtg agcaggcacc
70141 tgtcacaggt catcatgctg gatggcccct gcttgtcact caattctcag gcctgactca
70201 tagcacagac ccagagccct tgaaggaaag ggaggccagg agtctgcaaa aggccccgtc
70261 atttcccata agtgcagcct gtgagcgttc ctcctacctg ttctccctgg acctggagcg
70321 tcccctcaga tcacagggga cggcttaaag gaaacccatc cccaggactg ttggatccca
70381 tccagactga caccaattcc tgggacccat aatgacccag agtgccactg tggtgttcca
70441 gtcagagggg gggttacagg gaacaagtaa tgaatgaata tctgatttca gtttactttt
70501 caataagtct atcacctggt tattattttt taaattcctg ggtggataac tggaatgcat
70561 attcttggaa gcttcagatt gttcacgtag acccacggat gatgactatt tcagtgacag
70621 acgccggtgg aagccgttgg ttttacatac ctgcctaaat attttaatca gagtcaacag
70681 caaattcatg gaggagttag agagattgct gccatcatca aggtttaaag ggagcagaag
70741 cagtgattcc tgtgactcac ctgttcacct cttcaatgca gccagtcgtc agcagggaaa
70801 caactgtttg tcttccaaca gcagaacaaa gcaggaagct catcagtRga agctgccatt
70861 cctaaggtgt tcttgttaat gaagcaaata aacacagcct gcagctctca gtctctatca
70921 atcaggcttc ttggtcaagc agaggtacca tctctggttt ttaagggcag aaaacaacat
70981 gaaattcact ggaatcatat cagaagagcc ccacctgcag aggaggctgg agtttggctt
71041 ggaagataaa aagaagccaa gacagtgaca cggtatcaag gctgcgagca caagagccct
71101 gctttgtgag gagtgacacg gccacagcgc tgccccgcca cctgcggaca tcatctctgc
71161 aggtgcagga ctcggtcgcc ctcttctcta ttctcctcct gggcagtcct gctgctctca
71221 ctgagaatgg gttctaagca ccctgggctt tgcctcactt gctccagttt cagaatccca
71281 tgtggaagca gccaattagc cagcctagct cttggacagg agccctggat gccagagtgg
71341 gggtagggag tgggagaatc tacttgcctt agaatcacag agggaaatgg gcccagcctc
71401 tctctatgac tacacatgct ggcaaacScc caaagtagga agaggaaaat agatgcggaa
71461 gagttccctg ctgtgaaaca agtttgttcc atctcttggt tctaaatatt gacatctttg
71521 attgtacaag gcagacccct aactctaatt tgYgaaaact aaatagagat acataactca
71581 cagaccatga gggctgaatg caggtgctta cactgagtta tgagtactca acRtctctca
71641 aattctgctg tgctttcctc ctggcccacc tgctcccacc aggcctggcc tgctgtggaa
71701 cacctctcag ggatggttcc tgaagagcag tggagcacac ctcacagtcc aggctgcctg
71761 gaaggagMRg tgtcctcggg aaggatctgc cttgtttgct gggtaacagt tgaggatttg
71821 gctgtaagtt cagcatttta gaaggaatac gtctggaggc ttggtgacaa agatatttag
71881 ggaagaggta tacagataaa gattttagaa tgagcccaga gcaaaaaaaa gaaagagaaa
71941 aaaactaccc ctgtcccatg tgaatactga ctagaaagtc tccactctca agagactcat
72001 atacagggaa gaagtcactt ttgtggctgc aagtctcctc tgttcccagg gcggcccggt
72061 attttcagaa tgggctccat aaccaggtgc gtgatgtgga gctcggtgta accacgcaac
72121 acggatccct cctcaaggct gagcacttgg catcgtccag tgaatctcaa ccccggaagt
72181 cacagcctta gagagcatgg aaatggcccg aaggggactc agttccatgt cagaggctgg
72241 gaactgcctt taggatgtgc tatgttttga attcataccc aaaatttacc tcatttcctg
72301 tgatagccag aatttatgga gctggacagt aagaggttaa gatgagggtg atagctctat
72361 ggttacaaac aataagtcac tttcaaaaat cttgtttaca atacccctaa ctcagagact
72421 tttaagtcta gaagtttcag tatctaaggt gggaatgcct cctgcaaggc tgacaaaagt
72481 ggtttcagag gtgcccatag ctggcttctc atgctccctg atgtagggat agcaatgacg
72541 tcactgggtt agatataatc atggatcatg atcatccagg gaagctcatg ttgttactac
72601 acagcaggac atgaggaatc gagggatccc ctaggattgc tccttgtaat actctaacaa
72661 gttgtaaagg tgaataggaa atttcaccac cttactgagg ccagaccatc aaggagctta
72721 gaccccacaa gaatgacttc gcccataagt aagaggaaca ttgagtggaa cagtcaaaga
72781 cttgtgacgg tggcagcaat ggtcagtggt ggttgtagcc ttgaccgcac ctctctgttt
72841 cctacaaata tacacacaga aaatgctgag ataggagctg ccaatggacc aagagaaaaa
72901 ttgacagtgt gccgaggcct tcctcctgtg acataacttc agcatgcaga agagggcgat
72961 ggtgtgtgtg gctctctttg gagaagaagc aagtgcattt gtctttgtcg ttcggatggc
73021 tggattcata acaggtaaRg agagtttttt gtttgttttt cgagacggag tttcgctctt
73081 tcaccaggct ggactgcagt ggcgtgatct cagctcgctg caagctctgc ctcccggatt
73141 caagcaattc tcctgcttca gcctcctgag tagctgggat tacaggcatg tatcaccatg
73201 cctggctaat ttttgtattt ttagtagaga cggggtttca tcatgttggc catgctggtc
73261 ttgaactcct gaactcaaat gatccgcccg ctgcaaccac gcgaagtgct gggattacag
73321 gcgtRagcca ctgtgcccag tcgaaaaggg ttctttttgt ttgtatgaga agaagggtac
73381 gtgtgtgtgg tagcaagcag ttgactatat aMaataagca caaattcagg tgactgtttg

73441 gccagaggct tcccattagg gtttgtgtct ctgttactca tcgctgttcc tgataccata
73501 gcctgactga ttggttcaag aacagtcact gacccaagct atcctcYcat ttgagaatgt
73561 gaaactttaa tcagagacac taggaatcaa agtgtgtgga gctaagccag gtgcaggcag
73621 agccctgaaa gggcaaacca tatattcctg atgttggtaa tggtctgctc ctgtccttcc
73681 aaattctgta gatttgttc ttccttccat cctctgacct gctctagttt tcatccaatg
73741 aatcttctat tttgtttaag acagatagWt tctgtttctg ttgcttgcaa ttgaaaaatt
73801 tttaaRcaat actagcagta cagtggaatt gcatcaaatg caggtagcct gatggcattt
73861 caactctaag tgtcagggca aatgctggct ggatcactga agaagtaaat tactcagtgc
73921 cagaaacatt caaagcttaa gtgaaatgta tgaaaagaat gggaattatc caggaattga
73981 tgaacaattg acagacaaat gaaacacaat gttctggcgt taggcaagtc ggtggccgtt
74041 gctctgttga cttcttcagg aagaaataga tgaatcaact gaaggtcagg gaagccttgg
74101 ctgatggaga gttacaggga ctgaggcaag tgttctgcag gatgctacat aaaaaaggac
74161 taagtaggcc ctgggtgcac tttacgagtg acttaggaga ggttctgggg tgattttgtg
74221 cattctgtat ctcttgcccc tgctctgggc tccagctgag gcgcccagca gttctatatt
74281 ctatgcatgg aaatacagat aattaatgtc cactggggca aggccatggg ggacaggagc
74341 agtggttaca ggcctccctc tctcctccct gtagaggaac attctgaaga cgctccttga
74401 agactcctca ggaggtctcc agtcgcccac atgggtgacc ctatcaagaa tttctcccgg
74461 catcagYctt cYRtccttct cYgtgtcact cctMccagta ccccagtcct gctcctagaa
74521 tcacttccca aacaaatcct ctgcctggga gccttgccta gggctctccc ttcaggggaa
74581 ctcaggccac aRcagcctac attagacagc atgaaggctc tgggcttttg tgctaagagt
74641 aataggaagc cattgaagac atggaagccK gKgtttgatg aaatctgtgt ttttatgggt
74701 caacttggcc caaKgtggag gaWgagctgg cctagcggca agggctgtRt atcagcgaag
74761 accattttgg aagcttttgc aggagtctgg gcgaagtgag aaggctgact ggatgagggt
74821 gtttgcagac tggatgaggg tatttgctgt ctggctgtct Stctctgtcc ctctatgtct
74881 catggatttt cYYtcaacct acatacaggc ccttctgatc ttagcctcat gtgacccatg
74941 tcaatcctgg gccctggctg ctagatcagg accacgcctg tactagagcc aagccagaag
75001 tctacaagtt ttccctgagt gcttcagggg ccactgggca ctgggttcgc agagaaatga
75061 gcagtctgga gagtaagcag aaggggccat ggaggcctag ctccctgaga gcctggaaga
75121 agaatcatga ggaagagaag tcaggcaccc agagcctagg ggagcccata tatcaRcaca
75181 caaaggctca gtcatcagtc agtgttgctt gaacacaagt ctctagaaca caagagaaga
75241 gggccaggtg gatgtgagat atcatcttgt ccagcccccc tttcccagct gatgaggaaa
75301 ctgaggctca gagagatggg gtctcttgac tgcaggctgg ttcatctcct caatgcacca
75361 caacattcct tgacttgttc ctggcatctg gttttgcttt tatttttctt cacacatcaa
75421 cgtttccaat atagcatttg gagtttcatt gaaatgtcat cggtccctca gtgtacgctg
75481 gccccaagca gggccgaccc catgtcaagt cctcttgctg tctgctgaca tcaccacaat
75541 cacctttacc actgttagga caaaaacttg agacaggaac ttgccctcaa agcgggaaag
75601 gagccatgag accaaggaat gacttaaaca agtccagctt ggcgagtagc taagtttatt
75661 aggacttaca cacagggcac tcagcaagac agctcgagag atccRgcctc ccccaatgtc
75721 taaactgctt ttcagttaat tttctggctc tttgtctact atatatgagc aacgagactt
75781 ttacttggtt ggttctcaga tactctccag gatgtttgga ttctcagaga cacctggtcc
75841 tcagctgggc accatggaaa tggctcatta cccggccttc agggttcaag caggggacat
75901 gcacccttaa ataacctaaa ggggacctgt cacactacaa ccaccaccac catcatcaag
75961 aagccactgg ctgactcaga tacaccccYg ggaggacaag ggagagtgga tgctggtaag
76021 acagggYgag agaccatcac cagggaagga ttccattctt ggaaggacat ccaaccgggg
76081 ggcgggtctt tagtggagcc gctgtttctt ctcctgtatc caacagttct aactgtggct
76141 ttctccattt tcagctcttt cttttcctgg tcttctcatt gcttgttcct acgtgcctcc
76201 ctctcttcct cccccaatat actctttagt ctagagtaaa ctgcttcttt ccattcccca
76261 cactctccag tcccctctcc tcccttattc caggccccag catttctgcc ttctccttgg
76321 cgcacttgcc atcttgagta acccctcccc tgctgtgctc agctacagat gccaaagttc
76381 ctaacactga gttccataaa ctttacctcg ccctctttat tcccctaaaa aatgtctgcg
76441 ttttgtcccg gcctgtgtct ggctcacctt ccctggtggc tgcactgctc tggggtcatt
76501 gggtatcggg gaacctgccc cgatggtcac gtaggttctt ttctattttc cctaagtgtt
76561 ggccggtttg agaaataaag gcacagagta caaaagagag aaattttaaa gctaggtgtc
76621 caggggagac atcacacatc ggtaggttcc gtgatgcctc acaagccgca aaaccagcaa
76681 gtttttatta gggactttca aaaggggagg gagtgtatga atagggtgtg ggtgtgggtc
76741 acaaagatca cctacttcac aaggtaatag aatatcacaa ggcaagtgga ggcagggcga
76801 gatcacagga acacaggacc ggggagaaat taaaattgct aatgaggttt cgggcaccat
76861 tgtcattgat aacatcttat caggagacag ggttttgaga gcaaccggtc tgaccaaaat
76921 ttattaggcg ggaatttcct cttcctaata agcctgggag cgctatggga gactggggtt
76981 tatttcatcc ctacagtctc gaccatagaa gatggccaca cccaaggagg ccatttatag
77041 acccaccctc aggggtgcat tctctttctc agggatgttc cttgctgaga aaaagaattc
77101 agcgatattt ctcccatttg cttttgaaag aagagaaata tgactctgtt ccaccaggct
77161 cactgttggt cagagtttaa ggttatctct cttattccct gaacaattgc tgttatcctg
77221 ttcttttttc aaggtgccca gatttcatat tgttcaaaca cacatgctct acaatttgtg

77281 cagttaacgc aattatcaca tggtcctgag gtgacataca tcctcctcag ctgacaggat
77341 taggagatta aagtaaagac aggcataaga aatcacaagg gtattgattg gggaagtgat
77401 cagtgtccat gaaatcttca caatttgtgt ttagagattg cggtaaagac aggcatagga
77461 aattataaaa gtattaattt ggggaactaa taaatgtcca tgaaatcttc acaatccacg
77521 ttcttctgcc atggcttcag ccagtccctc cgtttggggt ccctgacttc ccgcaacaat
77581 tgggcctagc ctgcttttag actctgatgg atatgggtga gctcattgag attctcctcc
77641 agctcctgag cccactgcct cagcgactca gcagactcgg attttgcccc cttgtgcaag
77701 tccagtgagt cttgcacaag gttgactaca tccagcacaa caaggacacc tgaagtggcc
77761 ttgcccaggt tccgggctac ttttctcact atccgggtgg gggccccacg tgttctcagt
77821 gtctcaacaa catttatagg tacatatcgc caagcaatca aagggcgtcc aacagtggct
77881 ttagatctcc tgagtgtatg gacatcattc gcaatcattt ttgtggcttc gtcaaaatca
77941 agtgcaaaRg aaagcacatt gggtgtgatg tcaYgcagaa tgtcccttaa tgcctccaat
78001 tggtcagtgc tggttgcagt cagcctgctg gctgtgagtt ctgctgacct tgtgtatgtg
78061 ttctccacga tgctggaggc gatcccagcc gtggcagatg ctattcccag ccctacccca
78121 gctgcagtaa tgctcaggct cagccctgct gtaaatggtg ccaacaYaac gccaatgaca
78181 gacaggatgc cagtggagcc agacaccaca ttggcgatga cgcagcctct gtggaccttt
78241 tcaatctcat ttgcaatgac acgaagcctt tctatggact cctgaatctt ccatctgatt
78301 tgaggaaacN ttcaaaaacc actccctaaa ctgctgttct ttttgctgca tgtctttgtc
78361 ctcaatagcc acatatggtg taagattctt cagagcttca tagagagcat ctgcYtcttc
78421 cctgtaccaa taaaggacag atgattaaga aaggcagctt acttacctgt aaaatgagct
78481 caataagatc tattctgcat aaatcacaga gctgtttcta taagtcaaat ggaagtaaag
78541 ttctaaaaga taaataattc ttttcaaggc cgggYgcggt ggctcatgcY tgtaatccca
78601 gcactttggg aggccgagat gggtggatca caaggtcagg agatcgagac catcctagct
78661 aacacggtga aaccccgtct ctactaaaaa tacaaaaaat tagccaggYg tggtggtgag
78721 tgcctgtagt cccagctact tgggaggctg aggcaggaga atggcatgaa tccaggaggt
78781 ggagcttgca gtgagccgag atcgcgccac tgcactccag cctgggcgac agagcgagac
78841 tccatctcaa aaaaaaataa taataataat tctaYtcaag gttaaatatg ttttgtacaa
78901 catagatatg ccacagtact tattcagtat gggaaaaaat aagttaaagt catcttccga
78961 tgctctttag tggctatgtg tttgataaag atggctcctc atgcaaggga ggtcaatgcc
79021 atcactacca gggtgtgtgg agaggggaaa ctttgtagaa aaacaaagag aagtgggttc
79081 agagcaaaaa gaccttcatt cagaggatca tagaacagtg atgaattaac tttattagaa
79141 gagagcatat ccctttgtcc tcggcccctg gaaagggatc tctggggccc tggaatgtcc
79201 tgcctggtag gaacgtcttt gtttccctgg tggtttggct acaggccagt atagcaatgt
79261 gatgggtgat gggcgtttag ggctattagg tgtctcatct gccttccaga ggaactgagg
79321 actaaagggg ttagaccttg gggaagggct ggagactcca tgtctgccat gacggcagct
79381 tgtgatccag ccccaggaag agctctgcat tgtcatgagt taagaaagac ttactccttg
79441 gccaaacttc agttgggcgg ttctgagcct cttctcctct agacattgac cttggccttc
79501 catgcccatc ccgtgcttgc tgggcctgca ccgccccagc tcagcaagaa tcccccaaag
79561 tcaatttaga gaggatcccc ccagcttgct gtctgagcac acttgctctc tgatcacggt
79621 cttcacccat tcatttgata agcgactgtc tgacctgcct ttagggagca tcctgttagg
79681 ccagtttagc aggaaccccc cactctggtg tctcctgtca gtcctgttcc atctgccgcc
79741 ccttaccctg gtgttggctg tgactccccc tctgacttta ctgtattggg agctgggctc
79801 agtctccctc ccctgtaaca acaatcctga ataaaggctt ccttcctgat ttaatgtgat
79861 tcaYaggaat ttttccttaa ctgggtgagc ttccctggtt ggcaacaccc catgtgtttt
79921 gttccacatt gacacaggag aacacagcct cctgaggaca atgaagcttc acaactgaaa
79981 ccctctcaga ctctgccctg tatgtctctt ccttaatctg tgttctttcc ctgtgataaa
80041 ccacaatcgt gcaataacag ttatcagtga gttacgtgag tccttgtagt gaattatcca
80101 acctgagggg ggcttaagaa agcccctcaa gtttgcaact ggtgtcagaa gtaagggtgc
80161 actcgaacgt tgcagttcac ctaaatctga gtccttaaga attcaaacac aacctcagac
80221 tagaaggtgt caatttgcct ggcatgttct ctcttccaat gcctcatgcc ctttgacgtt
80281 ctgactaaag gccactgggt ccagccactg ccacctatga ctgaaaagtg gaagacctgg
80341 ctattagtcc tggctgcaca caaactctca agccacctta ctgcctcctg acctggcacc
80401 tccaaacttc agggtctcac tctcacatgt caccagccaa actaatgcct tagcaggtga
80461 cagcttcatg tgtatacagg tcaaggccct gtacaatttc atggaagctc ttgaggactc
80521 atgaggactt tgaaaatcat ctctataaat cactaatgct gaaaagacct gcatttggac
80581 aagatggtag agaaagtgtt tttctctttg gttggggact tgtgagtctt ccgagaaccc
80641 ttcaaattac accagttgaa gccactcgtc cccactctgg aagttggcct gaagaaacat
80701 tgcctctact ctatgtagaa cagaagaagg tttcaaatgg cagcccagaa ctccccattt
80761 cccaggcatg gtgtcgggga aacagtcacc catggcccac tatgcccatc agtgatggga
80821 tcaaatcagg gagggaagaa atatccctct cctactcacg ggtcttctgc aagtccatga
80881 ccatgttaca tttcatgctc tcagtgtctc ccttctctca ccctcctccc agccttgatg
80941 ctccagcctt cctgctgctc ctccaatcta ggtccacccc tgccccatcc tgacccctgg
81001 ctctgcccct ggcaatgtct ctctgggtct ttgggcggca tcatgctatt tggttcccca
81061 cctttgacct ggagttcatt ctaaggcctt caccacactc ccatctgatt gaattactaa

81121 tagcattgga cagacggcag gtcccagtct ccaagacata aaccaacagc cccagagctg
81181 tgcagggaca gagccaggat ccaccccagc tctgtccaac tccattccag cctctctgtg
81241 tcctgggtca gcagaaggat tctgctgtgt ttgtgcagtt cctcagaagt tcaggaaaag
81301 ggccccagac agcatYtccR ggaacacaca ttgatcaggt caagattttt ctgacttttc
81361 ccaaagaaag tcctgctggg cttccatctc taagaaaaag acttttggYc ctggtctctt
81421 gccagcaagt tttgtaaaag gtaacgctaa ctgagcacct actgtgtgtc agtcattatg
81481 ctaaatactt tacatgcatt atcttatttt tctcacaata aagctcaatg gatgaagaaa
81541 ggaaagcaga aagtagtcaa gtagcttgca aagctcccat caccagtaga tggcagaact
81601 gaccctgccc ttgtgggctt cttctcccct cagcctgagg tcactcactg ccagcgaagc
81661 acttggagga aatattcttc ctgcactgct gagagggaca tcctcaagcc cagcagaggg
81721 ggctgcctgg aggaggYgtg cctgccagag aaaactagcc cggggagatc tgggtggcat
81781 caccggggtg ccccaaggag gtaaccccat ggaggttacc tgggcaattc agccacacgc
81841 acRaatctct tccaggcttc atcgctagtc agcaggattt tcagatgcac tgggctaact
81901 ttcttctgga agtattcaat gacttcttca gtgaagcgtt tcttttctag ttggaaacaa
81961 aaaggataag attggaagaa agtttgctac cacataaatg gcattgagta taaggtggtt
82021 cggtgttaat cctcctgaac cagctgtcac atggggtatt tttgatggag gcaccagtgc
82081 tatagactga attgtgtgcc cYccaaattc aaatactgag gctctaaacc ccagtgaccg
82141 tatttggaga ttgagccttt aaggaagtca ttaatttaaa gggagtcata agagtaggga
82201 cctactccta taggactggt gtccaattaa aagaaaaaaa agagacagat gatctctctc
82261 tctctctctc tctctctctc tctctcacac acacacacac acacacacac acacacacac
82321 acacaccact gtgaaggcca tgtgaggaca cagccctgag gcagcctctg taagccagga
82381 gcagagccct ccccagacat caaccctgct ggcaccttca ccttggactt cttgcctcta
82441 gaactgtgag aaaatacgtc agttgttgaa gccacccagc ttgtggcatt ttactaaggc
82501 atcctgagaa gagtgattca ccagggaagt gccacgatgc tttgtggagg aaccaatcta
82561 tttcagctga gaatcaccag aaagtgagct ttccaccatg ttttctcccc atgtacggga
82621 aatattccag tgatcgcttc ctcctgccat gtgcctattg tcaaacactt tacaccagtg
82681 tcttcaaaat ctgaattttg acccatcagt gggttatgac aatagttcat tgtgtcaaga
82741 gcagcatttt aaaaagaaaa gagcattgaa aatatcagag tatattgaac atagcaagag
82801 tgagtattct tttgtaagat acttgaaata taYctataga cccttatgta aattcctggg
82861 atctagatgt gtttaggatg gcagcatttt caatcttcta gaagggacag agggtgcata
82921 tggtttatca ctgaccactc ccagcaggga ctgggcccca ccttgtaatc aaaggtacaa
82981 atgtttcaac caggaaacca atgaatattc accttaagtg gcatacataa tattatcact
83041 acaaagcatc ttagtacaca gcaggattac tgcccagatg agttgcccca aaaatgtgtg
83101 gtttgcaggg atttgggaat tatggaactc cagagaaggg gctgggaacc agtttgtatt
83161 tatgtatgca ctgggacatg atgtataatt gtttcctgct gtgctcatac gtgcgaggcc
83221 tcattctcat ttcacagaag gagaaattga gacccagaca gcgagagaca ccacccaacc
83281 cagcaagcag gaagctcagg atttgactca aggccagcgt tcccttgcca ttactttgaa
83341 gactccatat atttaaaaaa tgcctgtggt ttctgcattc tcccctgtag tatacctgtg
83401 gggttttaaa cattttattt gcacagagga ttcctcctgc tggtcatggg ctctctcaga
83461 Ygggcagtgg agggggttct ttaggtgggc aggggagtgg gaccctcagc actgatcctt
83521 gtggcccctc cgtgttgcct ctgtcccctg caggctggtc tcagggtaca cggctcccca
83581 ctggcacctc tgaacccagc tggccccgtt agtaaaccac agaYcatctt gtttcagagg
83641 gagagcttct tccttggccS cccagcccag cccttcctaa ggattctggg attggctgag
83701 WgctctcRgc ttctaatctt ccctgccctc ccctcatctt ggacaggact cagcacagcc
83761 cagtttgctg tccagggcgg cYcctgagca gagctgactt ggttgccatg gcaactctgt
83821 aaggggagga agccctcctg aggaaatggg gcttagggaa atgaacgcag tgctgacttg
83881 gctgggcgag ggattgtctt gttttttctt tttccttctt caccaaaagg aggaaggaaa
83941 ctcagggttg gatggaggga acctccagca ctaaatttgt gttttattga tggggtgggg
84001 atggtgaata ttaaagcaaa attaaaataa aaaccagacc tgaaacattc ccgggcaaac
84061 aaaaccaacc agaccttcag aacagctgta acccccсatg aaactgcaag ctatgaatct
84121 gaactggggt catatcagat gggcgcttct gtcagacaca aacacagctt aacctcggcc
84181 agtcatgagc agccagctga cagagggcca tgctgagatg tggatactgg gaggtcacag
84241 aaagaacaag aactccaggc agcaggttca cataatggac acagaaggag ctgctgtaac
84301 cagctgtagg gacaggggga ggggatgggc ccagaggacc ctgacaaaca ggaagtggcc
84361 aatctggcta agaccagtaa gatctaactc agcactgggt ctgacccacc tttcacccca
84421 gactcaatca taccctcctt gtcccacacc cacctgtgcc acagaagttc catgcttgca
84481 acaaaactgc gtggcccccc agttccaaga aatccccgcc tcttcccata aatccttatg
84541 cttactctgc cccttaactt agtaaaccca caaagtcaga aaccccaaac cccactgggc
84601 ccgactggct cttctgagtc cgcctgcact cccacccctg agcctgcgct tctgctttgc
84661 aataaaagct aatttgcttt ggctttgcct caaattcttt ctctcttcgt tgtcagaatc
84721 ctggacaagg ctaaggtccg gtctcacctg cattcagaga cctccctaaa cccaccggca
84781 tcaacatgac caggaatttt ccaacaaggt aaagcaaaac aggcagtttt atcactggaa
84841 ctaatcaaat aatgtcctaa ttgcacttcc atattcaccc gataaacact tgtcttccac
84901 agtttttcat cataacacaa aacctcctcc agtttggttt tctctattta tgaactcctt

84961 cttattcaga taaatgttgt tgtgcctcag atgtttcttt gacaaggggt actttcgatg
85021 attggactca gcccagataa agaggtgctt gtgacaggca gtgcagggtg ggtgccctaa
85081 cacagctcgg gaagggggtg gcttccactt acaatcaaca ggggtcaggt agaggggatg
85141 ggaaccaccc cttagggagg ccgtactgtg tggcaggcac tggctggccc aggtacccat
85201 gtaatgtttg atcctcatga caacatgggc aaatatccta gtggtgcagg tgaggaggag
85261 taaggagctc caggtcacac agaggggaca tggccaagac cagaatgggg tccagttcag
85321 tgggaagctg ggtcactcta cccacagggc cactcagagc agaggggctg gtgcagggca
85381 agagctggag ggaggggacc ggcctctgct ggaaacaccc cagggtcagc caggagcaga
85441 gagggaggtg agaggtcagg gctgctggtc aacctcctct cccacccctc cgctactcct
85501 ggccaaccct caaaagccaa ccctgctgtt tccatttccc caaggctgtc ctgccccact
85561 taagtcattt gctaggtgtc agggtaggga ggaagtgggc acctgggcca tgacccgctg
85621 aggagaggtc aggatgcaga tgccccccag aaccctgggt caggggactc cacgtggcct
85681 ctcttggggc ttggggcagc ctcatcRgcc acctccRtct gggttccgtt ggggctcact
85741 cagccttgaa gacaccaccc tccattctag ctgtgggtag gaaccagcca gggaagagga
85801 ggcgagccta ccagcagcca ggtctctgaa aggcttacct tggaactgtc cagccactgt
85861 ccagcctgga tggttttgct gcacccttga ggaagagaaa acaaattgtg ggattgaatg
85921 tgagccacct atgggcaaag ggaggcttga acagctgggc ctctgcagat ccctctgagg
85981 tggcagcaaa tgccaagacc aacctaaccc agattctttc tctatcacca agtccccatc
86041 atctgctatt tacagacact caactcattc cagctccctc ttccctcact ctcacaccaa
86101 ggaaaagtcc gcttgtcctg ttggagaatg aggagaagat aatcagagga ggggcagcca
86161 tctgatcatt acagaaacta cacagtctat acgcagaaat agagatggga aggaacgttc
86221 taacaatgac cggaaacatg tgtgacaact aattgatgat agccagtctg tattgagtgt
86281 ctgtgatgtg cagggcttct tttttatttt ttatttttat ttatttatat tttttgagac
86341 ggagtctcgc tctgtcgtcc aggtggagtg cagtggcgcg atctcggctc actgcaagcg
86401 ccgcctccca ggttcacgcc attctcttgc ctcagcctct cgagtagctg ggactacagg
86461 tgcccgtcgc cacattcggc taattYtttg tattttcagt agagaagggg tttcaccgtg
86521 ttagccagga tggtctcgat ctcctgacct catgatcagc ccgtctcggc ctcccaaagt
86581 gctgggatta caggcgtgag cggtcgcgcc gggcctgtga tgtgcagggc ttctgacatg
86641 tgtctcattc aatgatcaca cacagcccgg gagatgggta ccactttccc atttgaacac
86701 agaggaatcc acaaaagtta gggaacttgc cagtgttttc tattgctaga gccaggactt
86761 ctgcttcaac ctcatgtctt caatgactgt gctctactcc cttaaatata caatattgag
86821 tcatgtatac acRctcatgg taaatgactc ctcattgcaa atccatgtct cacactccgt
86881 gctcaacttt tcgtcttggt gaattcactg gtgctgcagt ggacagcctg tgtccccctt
86941 gcttctctaa gatcctttgt cccacaaaga cctggatcct gccagatgcc ggcctagccc
87001 ctgctgtggc tgccgagccc tcctggcgct gccctgttcc ctctgctgtc ctcccctagc
87061 tgatggacct tagggtgttt ctaattcttc atgtaaatta tgctacaaag aacaYcctcc
87121 ttcacgtaaa actgcccaca tttgatttta tttccttaag atatttgtag gagtagaatg
87181 actaggtcag agggtatcaa catctcaaaa tgttttatac ctagagccac actRcatgca
87241 aattgtacca attcacactc aaactagcag tatatgaaga tgcccattct ttcagtctta
87301 ctggctctca gtagtctctt tatcaggaca aaaaatatat acacaacatg taaatttgag
87361 aactaaatat cttgttatca ataccaatta gtacaaactt taaggcacga aggtagcagt
87421 tgagatgtga agcactgaac aaaggagaag gaggattaat attgtcattt tttaaagaKg
87481 gaagttaaga aatactgggt acaactgagg gtataagaaa gttgtaagta aattgtttac
87541 agctacaaag ttgctaatgg Ragacataaa actggtgata gaatgttagg aaaaggtgga
87601 gagRagagat agagggaggg tacaaataag atgaaaatga ctaataatga tactataaat
87661 aataacaata ccaYttcagg cccatcccag ttaggggcgg ctccaagcat taccattgtt
87721 cttttctcac cgtgtgtctc tgcgcactcc ctggcacacc cacttccagc agctggtctt
87781 tgtttgcaga ctttatttta acagctcagt ggttcccggg ccccctgatg gggttccatc
87841 ctccagctct catactaaag ccccctcctc aaaagccact cgcctccctc tttccctttc
87901 ctgcctttgt ctcagggttc agacacacag atgtctgtct tctggggccc ccatgcaRca
87961 agggtaaaag ggggacccag tcctgggcag cNttcaccag gggagtatgc agaggggcRt
88021 ctgggcgctt cttgctgtaa gggacagggg tgacaggaga ggggcatccc ttcScttSt
88081 tcctgctttg ttcactgtgt gaccccctag gccaRggSaR ccatctcaac accaNgagcc
88141 aKKagaacNg ggaggaccag gagagagcag aggagcagga gggcagagag ctggggcctc
88201 ggactcaccc gacgcttgtg atgagctgca cccaggatcc catcctcctt ggtcattgtt
88261 ggcctggctc agacgctgat ctggggcctc tgctgaatgt tgaggctgga tactgactgt
88321 tagcctcaac taggacacaR tgcKtcccct ctaggcgagt ctaccagttg gtcaattccg
88381 ggaagtgatt ttggttccta gaaaaaccca catggctcct ggcccagccc aagccctgcc
88441 caattgtgca gcccagacag ggagccctcc ctcccacctc agggctctga gccaagctca
88501 ccagatgcag aSgacaaaga Ytttcagcaa agcagctccc tccatgtcgc tgcggggcct
88561 cctccttggg caggaaaaga ggggttgaga cagtgtgctc ctcctgagaa atcaccattt
88621 tcatttcatt tttaataaac cgttttgctS tgtcacccag gctgcagggc tttgacgcca
88681 tcctagctca ttgctgccta gatctgctgg gctcaagtga tcttcctcct tcaaccttct
88741 gagtagctgg gactacaggt gcacgctacc atgctgggct aatttgttaa tttttgattt

88801 catagagaca gcatctcctg tgtttcYcag gctgtcttga actcctggcc tcaagtgatc
88861 ttcccgcctc tgcctcccac agcactggga ttacaggctt cagctatagt gctttctatc
88921 ttgatgggca tctcttttg gtgaagaaaa atcaggccgg gagcggtggc tcacgcctgt
88981 aatcccagca ctttgggagg ccgaggcagg cagatcacaa ggtcaggaga tggagaccat
89041 cctggctaac atggtgaaac cccatctcta ctaaaaatac aaaaattagc cgagtgtagt
89101 ggcgggcgcc tgtagtccca gctactcggg aggctgaggc aggagaatgg tgtgaacccg
89161 ggaggcggag cttgcagtga gcagagattg cgccactgca ctccagcctg ggcaacagag
89221 cgagactccR tctcaRaaaa aagaaaaaaa aaaaaagaaa aagaaaaata agaccacgac
89281 tcagatttca cagaaattac ccccctgatt tcatggactt ttggcagctc tatcaggttt
89341 ctgagagggg aaaatcttag gctgggaggg gaggtgactt gcccagagtc gcacagctct
89401 tggagagcag acgggaagcc aacatctgcc cttgtgactc tgctattgaa gaaaaacggg
89461 tgattttagt aatccaacag tgtccaggtt tgattctgct tctacaaaca cagagatcta
89521 agctctcact cctgtagtta cactcacaat tccacaggtg gcttttgaat ctcatgatta
89581 caaagctggc gtcatgccac aggcaccatt ctgcaatgcg accttctcac tcccctgtct
89641 cagccatgtt tcctattgcg ggatctggcc agcagccccc agtgcaacgg ggctctctct
89701 ttgctcctag gcggatcggc aggttgagaa ataatagaca cacacaagat agtgaaagct
89761 gggtccaggg gggtcaccgc cttctggtca acaatgcact ggatatacca gcatttatta
89821 ttcagtttag tgagggcggg ggtaggttag tgagggattt agggtcattt gattacgagg
89881 tgggatggtc acatggggat gaagtaattc tttaacataa cattcgtatg tagaagtaca
89941 gtgcatttgt atgtagaagt acagtataca gagataagaa tttacaatat agtgtgtgcg
90001 tcagtaattt ctaacagagc cttaaaacag aaacacaatc tttccataac ctatgattag
90061 caagatatta atcagcagta acaattgcaa caaaagctgg ttactgacaa tccatggaaa
90121 caggacgtga agctagataa ccggttagac tagaaattct cagaagggag tatacctgaa
90181 ccctaaagat gcctagaaga gccacggcaa gatgatagca tttatagccc tagcttgtcc
90241 atatggacag gcgcccccct gcatccattc ataggctctc tgcaggggga agcacatcac
90301 gtgctgttgg ttcgttctgg cagtccaacc tggcattgtc tttacacaat cctgcatgca
90361 attttgtatt tacaataatc aggagcattt catcttttat tccgtagcaa tagtttcagg
90421 gggtctccct acagtttccc attagggaca cagcacagcc tgtttcttct ccatcacccc
90481 agaggtctga acagagggac agaggtgagg gccacagaca tggagtccta tgatctacct
90541 gctcagcctc ctgacactgg accctgggca agtcactccc ctcccagggc ctcagtttcc
90601 ccacagggcc tggcacaagt aggtgctcag actcagggtg ggtgagatga cctctaggtt
90661 ccttgtatct caaaagaaat ctcttgcctc atcaggcctc agcagccgca ccccggggcc
90721 agcctggttg tggccagacc ataaggacag accRggctcc aggtcacctc cccacctcca
90781 ccttctttga ttccttcaaa tttcccaaag agatgggcac ccccaacaca aaccttcctt
90841 ctggggtcac ctggccatgt ctcagcagtc taaaacagag ctgagcccag agagctttgt
90901 gaacccatct gagctatttc cccacctctc tctacggttt aagggcccag caggagggag
90961 ggagcaatca gactcaagcc tggRtgcaaa tcccggctct accactgctt tcctgtctga
91021 tctgaacgag ttacctaacc tctccgagct tatctacaaa agctgaatga tccttccctc
91081 atagagctat tgcgagaata aggagatggR gggaggtcac accatcccca acttaccaag
91141 ggatcttcct ctgacagaga ctgagcaaga tccagctggt ctgagctgtg tggatctcac
91201 ctccagctgt gcacctatat aataaccaga cacgtcctcc agcccccaag atatacccag
91261 gaattcgaaa ggtaaaatga aagtcacaac ttcccagcag ctcacaatca agcacagcaa
91321 acacgctgct ccccagcacc tcctgcagtc cagccccacc ctccttgctg ctgcgcttag
91381 aggagcagcc tgagaccaga cctccaggtc tctttcatcc aacccacctg cctggcatcc
91441 tcggggttgg gggtctgcta tagtcttcag gaagaaagac ctgccactga catactgtgg
91501 gacccttcag attactcccc gtcctcccgt gcttccacat tcccctgatg cctgccttac
91561 tcctgcccat accatcagca tttgcacagt ctcctacaca tgccaagtgg atccttattg
91621 tatatttatt gtaaggccat tggggttatt ttggttttgt tctggagaca gggtctcact
91681 ctgtagccaa ggctgaagtg caccgccaca gtcatagtgc actgcagttt caaacccctg
91741 ggctcaagcc atcctcccac ctgagcctcc tgagtagctg tttctacagg catcagccaa
91801 ccacccagct aatttttaaa aaagttatat tagagatggt gtctcactat attgactggg
91861 ctactcttga actccaggcc tcaagtgatc ctcccacctt ggcatgagcc accgcaccag
91921 gcctgctctt gttgaaggcc atggcttgtg gcttctagat atttgcttat gtggggtctt
91981 ctgccgggat cccctgcttc caccacctgg aaaaatcatg ctcatYgttc agtgacaaaa
92041 tcaggcattg cctcctgtca gagatccccc agcatcctcc caagcagagg ccactgcccc
92101 gtgttccact cccgaggcac acagggtgga gggactagag accactgctc agccacactt
92161 ctatggctaa gaagtcatgc aaccagattg ctattggact tttgtgcttt gccccaggca
92221 gaagggcctc aaagacagct gaacctcaga agccaacctg acccttccat gggggcctat
92281 ttcatcctga gcaaacatcg ctgggggcat cgtggccacc tccccccacc catacgcaca
92341 cacaccagct tttgagacag gaacagaaag gtgattacac agggactgca caatacactt
92401 tgcaatacat tccatttgca aatgcttctc catttgattg aattttataa tcatcccctc
92461 aatgggccag gattatgatc taagttttca aatgaggaaa ctgaggcaag gggcgaggaa
92521 gttgctttca tccagcaact cagggaggag ccaagacttg tcaaccaaag agggtctggg
92581 acacgtctca gtcagtttag aggtttactc tgccaaggtt gaggatgcac ccaggaaaaa

92641 gacacgcaag tcacaggaag atctggggcc cactcattat tcccagaggg ttttgagggc
92701 ttcaacattt aaatgtgtaa aatcaggcag gcgtggaagg agggaaggaa aaaagacagg
92761 ataggtaatg agctgagtgg tcacattttt gtgaggcttt gatcggccct ccctgaatcc
92821 ataggttcca tgtgaaaggg ggttagagga acagtccatt attcattctt cttgtgcaca
92881 gttaatctac attttctata agataaaaat aaacagaaag tagaggaagt aggcatatgt
92941 gcatttgtct gagYgtgggc agaagggtca cttccactct tgtcattgtt aggtaccagt
93001 caagataagt tgccagtttg cattgtcaca gtgaaattat acaaagtctg tctcagggtg
93061 aaaatttggg gacctacaag gattttcttt gtgaggaatt tgtgaggcag gccaactggg
93121 gtatatgttg ccttccatct ttgcaaatat ctgtttagaa acaaaaacaa aggcaggttc
93181 tggagtgact ctgtttccag gcttaactgg ttccactggt atagtgaatt tggggccctg
93241 agatttgatt ttcctttcac atttctcttt tgtttttcaa tatcttttgg agaaaacagt
93301 gtagaagcaa aagaatctcc agtcacgggg ttggtctgaa ttctgtaaca caaagtattc
93361 caccttccgt ttgcNtgcag gggtttagaa aaaggacagt tttgatttct ggtgatgcca
93421 agtcagaaaa atgagagaaa aacaaaagga atatgttagt ttggagactt gtagccagga
93481 aagaattcaa tatctagtgc ggataaattg tagacaaata ataaagctgg aaaacaatgg
93541 acaaagctag aatctcataa cagacgtttt atagtttctt ttaaaacaca ttttctctct
93601 ctagtactct atttctacca aagacaaatc atagtaggac gcatttattt gcaaaataag
93661 ttttagtctt ctaatacttg gcctgattat ttgcataaag tgcagaagaa taatcttgtg
93721 gcataaggag tctttttaa aatttgactt tgctaggcca ttttcgtatt cataagctat
93781 ctcagagtag gccYtttaaa accttgagct cagccacgaa tttatgtttg cctgcagata
93841 cctgtatgaa ttgggtgagt tcctttcttc tcgaggcccc atgatcactt gggcttcctg
93901 ggcctgtcag aaagtgatat tctttattta ccacagttca ggaaccctgt aaaggaactc
93961 tataggcaaa ggacttaata aaggactgaa attaatttca gtcctttagt aattcagtcc
94021 catgtaatta ccttcttcca tttgatattg gcttagcaac cttcctgaac acatcagctt
94081 tttttttttt tttccagttt tgctgtcagg gcccagggga gggagaataa ggagttatgg
94141 cttgctggat acaaagtttc cattttgcaa gaagaaaagt tctgtggatg atagcggttg
94201 cacaacaatg tgaatgtaat tccactgaac tgtacaccta aaaatggtta aaatgatcaa
94261 ctttaggtta ggaatatttt accacaattt aatttttgtt tgtttgtttg tttgcttgtt
94321 tgttttgaga tggagtcttg ctctgtcacc caggctggag tgcagtggcg caatctcggc
94381 tcactgcaac ctccacctcc caggttcaag caattctcct gcctcatctt ccccagcagc
94441 tgggactaca ggtgcatgcc accacgctca gctaattttt tgtattttca gtagagacag
94501 ggtttcacca tgttagccag gatggtttca atctcctgat gttgtgatcc acctgccttg
94561 gcctcccaaa gtgctgagat gacaggagtg agccactgca ccaggcccaa ttaaaaaagt
94621 ttttaaaatc cttgaaaacc cacctttgca gaaaaaaaaa taatagcacc tgagcctttt
94681 ctggtttgct aaacattttt acacacataa tcatacaagt ttcaacaaca gccctgtgat
94741 gaaaacattt tcttcctagc ctcgtttccc agaagagaaa acaaaggttc aaatggggtg
94801 aaatgagtta ctcagaaaga ggcagagctg ggcttgtagc tggacctttc tggttcttgt
94861 tttgtttcat tttctggacc atcctgtttc ccctaaagga aatgatttcc tgaaatttct
94921 caagtattct gaaaatStaa ttttcttcct cactttctcc ctgttggctc aagatgaatc
94981 acaaaactcc aaataccatc agaatgtgtt gcccaagtcc cacctgttct atggctgatg
95041 gccacaggga gctgggtatg gctgggcYgg ttatgaccaa ctgtaacccc atcaaggatt
95101 aacatgctcc tgaaacagac cccatgtcca gaaatttgca aaacctaccc tgaaatgttg
95161 acatcataga gtgtgcccca atgtcctgtc cagatttctg tccctcccat gagctggaga
95221 ggatatttgg catcagacag ctggaaatca ctttccccag tttttggctt catgactcag
95281 ttttcttgct tgtagaataa agctaatcat agaatttacc tctcagaatt cctaagagaa
95341 taaattcagg cagggtgggg aatatgcatg gcatatagta gatgctcatt aaacaagaat
95401 caatatttct gagcactcac tctgtataca gggcagtgag caatcatgat gtctgcatta
95461 tctccatttg taactgaaga aaagttcagc tgctcactgc ttgcaaaggc aaatgaacaa
95521 ggatgagaag cagtgaaagg aaagtgactt cattccaaaa gcccatatct ggggatgagg
95581 tcaggctagt gccttagaga aaccatttca gattttgaac tgagggcagg tgtttcaaaa
95641 gggaaacttg atatgaaagg catgcaagag tggtgaagaa gtgccagtct acatgacttg
95701 tccaatgacc atcttgagtt attgccccat cgggtgtatg gactggggac atctccagga
95761 tagccaggtt gaaaattcaa tgtagccttg aggcaattct gtagcttcct ggatggtttc
95821 aagatttctt ctactggggg aacacactcc caacatttca atggaggttc tttctatttt
95881 ccataaatgt tggccagctg agaaataaag agagacagta caaagagagg aattttacag
95941 ctgggctgct gggggtgaca tcacatatca gcaggactgt gatgcccacc tgaatctcag
96001 atcagcaagt ttttattcag ggttttaaaa aggggagggt gtgtaagaac agggagtagg
96061 tacaaagatc acatgcttca aagggcaaaa agcagaacta ctaataaggg tctaacaaag
96121 atcacatgct tctgagggaa caggacaaag ggcaaaagca gaaatactga tatgttcagc
96181 ggtgcacRta ttgtcttgat aaacatctta aacaacagaa aacagggttt gagggtagag
96241 aaccggtctg accacaaatt taccagggtg gactttttcc ccaccctaat aagcctgagg
96301 gtactgcagg agaccagggc atatctcagt ccttatctca accgcataag acagacatcc
96361 ccagagcggc cgtttataga cctcccccca gaaatgcatt cctttcccag ggtattaata
96421 ttaatattcc ttgctgggaa aagaatttgg caatatcttc cctacttgcc catctgttta

```
96481   taggctctct gcaagaagaa aagtatagct ctttttgccc gaccctgcag gcagtcagac
96541   cttatggttg tcttcccttg ttccctaaaa tcgctgttat tctgttcttt ttcaaggtgg
96601   actgatttca tattgttcaa ccacacatgt tttacaatca atttgtacag ttcacataat
96661   tatcactgtg gtcctgaggt gacgtacatc ctcagcttag gaagataaca ggattaagag
96721   attaaagtaa agacaggcat aagaaattat aaaagtatta tttgggaact gataaatgtc
96781   catgaaatct tcacaattta tgttcctctg ctgcggctcc agccggtccc tccattcggg
96841   ctccctgact tcctgcaaca tcttccctgg aatttctaag caatctaaca tgattaaata
96901   aaatacagtg catgggaggg gagttcctag tagaaaagga gaacaaagat tcacacatta
96961   ctaaaagcaa aacaagaaat ggacaaagag aggagaccag aaacaaacaa acaaaaaata
97021   aaatgggata cttggttaca ttttcatctt ctccatgcac tctctagtac tcttgatgtt
97081   tctcctatcc cYatgtacag accagaaaca gactcagaaa ataaaatcat ttccccaagg
97141   ccacacaact ggccttggct tcctttttg ttagagacag gatctcactc tgttgcccag
97201   gctggggtgc agcagcagat caaacgtctg ggctcaagtg atccccccac tgcaggaacc
97261   caagcagctg aaactacagg catgtgccac catgtctggc taatttttta actttcttgg
97321   tagagatgag gtctcatgat gttgcccagg ctggtcttaa actcctagcc tcaagtaacc
97381   cttctgcttc agcctcccaa tgtactggag ttgtaatcga gtaacccagc attgaaatgt
97441   actttaacat tatttttctt tctgtcttaa tctcagaatg tagctttaaa atgtacttca
97501   aaattccttt cccctccctt tctcaccaag ctctcttgtg cacagtgctc acttatctaa
97561   tgatgtgggt gcatagaaat tcaaggagtg aattttgaaa aaaaaacagg cagagaaacc
97621   caggtgcaaa attctccggc ttagggggag ttaccaacaa ttaatccacc accaacaggg
97681   taaagtccaa atgatgccaa ttggacctcc tgacgggcta cttctcaaga cagccattgg
97741   aacaagacac agacctgcac cctcctgcac cactcccaca cacctccaat tccaattttc
97801   cctttctaaa cccctctctc cagcctaaag tagaaatgat cttttaagag catgagcctg
97861   atcattcccc cactgtgagc acttgaatga agctgctttc ccttcaccgc aactcacttc
97921   ctgtgttttt ggccactgag tggtgagcag ctggacttga gccagttata gcattacagg
97981   agggagccaa tgcacccagc atggcttccc tttttcacct actgatcctc acgattcagt
98041   gcttaaatta tagaaatttt ttcatttcta agacaaataa tgatagactc acaagacaaa
98101   acagcacagc tggaaaaata gttgcaaaac tagcacagag aggccctgca gactctttgc
98161   ccagtttccc ccaacggtcc tccagttcat ctgatgtagg agctggactg ggctggtctg
98221   gaaactcagt ctcagctgcc tgaagtggtc acttatccat gcgctccctt agaggggatg
98281   acaaggtcct aggagaagtc aacatcaaag ccacctccaa gctgtggtca gcccaaagga
98341   gatggtcaca aaggcagaat tcagaaggag gaagagactt agagccaagg gtggagatgg
98401   aacctagggg tgggccttga gttacttagg tcaccgatgc caagctgcca taacaaagag
98461   ctcccaatta ccgatatatt caatcaatga tataggagat tgctgaactg agacattcac
98521   accaacgttg ctaggcagtt tcagtggcag gtggttttgc tcccgaaggt cgaccagggg
98581   ttgccactca gagggagttg ttgttttcca cgtggaagct ggcgcttacc catacatccc
98641   agcccaaaga acagaagagg acacaaaaca caggaatctt cccccttaaga gaaacaccta
98701   caagttgcag acatcactgc tctcaccttc cattggccac aatttaggaa gaagacccca
98761   tggagttgcc agggatactg ggcagtcctg tgcctgtgga aactcagcag cYtgatgatc
98821   aaagacgaaa aaaaagagag gaaataaaaa cgggtaaaat gctcaatttt ataactttta
98881   aattttcatc ttttgccacc tggagaacct gaaatggaaa cgaaaagtca caacttctca
98941   gcagctcgtg atcacgcaaa cccactaaac acactgctcc ccggcaccac cttcagcccg
99001   gtcccaccct ccttgctgct gcacttacag gagcaacccc agacccagtg
```

[0242] The following cDNA sequences 1-6 show human cDNA structure for transcript variants 1-6 of APOL3. cDNA sequence 1 encodes the longest isoform (1), which is shown in amino acid sequence 1. cDNA sequence 2 differs in the coding region but maintains the reading frame compared to the variant in cDNA sequence 1. This variant encodes isoform 2, which is shorter and which contains a unique N-terminus compared to isoform 1. cDNA sequence 3 has multiple differences in the coding region but maintains the reading frame compared to the variant in cDNA sequence 1. This variant encodes isoform 2, which is shorter and which contains a unique N-terminus compared to isoform 1. cDNA sequence 4 differs in the coding region but maintains the reading frame compared to the variant in cDNA sequence 1. This variant encodes isoform 2 (amino acid sequence 2), which is shorter and which contains a unique N-terminus compared to isoform 1. cDNA sequence 5 differs in the 5' UTR and

coding region compared to the variant in cDNA sequence 1. This results in translation initiation from a downstream ATG and an isoform 3 (amino acid sequence 3) with a shorter N-terminus compared to isoform 1. cDNA sequence 6 differs in the 5' UTR and coding region compared to the variant in amino acid sequence 1. This results in translation initiation from a downstream ATG and an isoform (3) with a shorter N-terminus compared to isoform 1.

APOL3 cDNA SEQUENCE 1 (SEQ ID NO: 2)

NM_145640 Homo sapiens apolipoprotein L, 3 (APOL3), transcript variant alpha/d, mRNA

```
   1 agcaggaggg tgggaccaag ggtgctgctg gaccaaggat gggactgggc caagggtggg
  61 gctgggaagc atcctgtttt gcatgtttga tcaggagctg ctgccaagtt gtgactttca
 121 ctttcccttt tgggttccag ggtatatctc agagcctgga gaacgtgtct ggttattatg
 181 cagatgcacg gctggaggtg ggatccacac agctcagaac agctggatct tgctcacact
 241 ctttcaagag aagcttcctt gaaaagaaac gctttactga agaggccacc aaatacttcc
 301 gggagagagt cagcccagtg catctgcaaa tcctgctgac taacaatgaa gcctggaaga
 361 gattcgtgac tgcggctgaa ttgcccaggg atgaggcaga tgctctctac gaagctctga
 421 agaagcttag aacatatgca gctattgagg acgaatatgt gcagcagaaa gatgagcagt
 481 ttagggaatg gtttttgaaa gagtttcccc aagtcaagag gaagatccag gagtccatag
 541 aaaagcttcg tgcccttgca aatggtattg aagaggtcca cagaggctgc accatctcca
 601 atgtggtgtc cagctccact ggcgctgcct ctggcatcat gtcccttgct ggtcttgttt
 661 tggcaccatt tacagcaggg acgagtctgg cccttactgc agctggggta gggctgggag
 721 cagcgtctgc tgtgactggg atcaccacca gcatcgtgga gcactcatac acatcatcag
 781 cagaagctga agccagcagg ctgactgcaa ccagcattga ccgattgaag gtatttaagg
 841 aagttatgcg tgacatcaca cccaacttac tttcccttct taataattat tacgaagcca
 901 cacaaaccat tgggagtgaa atccgtgcca tcaggcaagc cagagccagg gcccgactcc
 961 ctgtgaccac ctggcgaatc tcagctggaa gtggtggtca agcagagaga acgattgcag
1021 gcaccacccg ggcagtgagc agaggagccc ggatcctgag tgcgaccact tcaggcatct
1081 tccttgcact ggatgtggtc aaccttgtat acgagtcaaa gcacttgcat gagggggcaa
1141 agtctgcatc tgctgaggag ctgaggcggc aggctcagga gctggaggag aatctaatgg
1201 agctcactca gatctatcag cgtctgaatc catgccatac ccactgaccc cagaccagtg
1261 cagccagcag gggaggtgag ccatacacag gccacgacaa aatgcaggca ttttattagg
1321 gggataaaga gggcaaggta aagtttatgg agctgagtgt tagtgacttt ggcatttctg
1381 tagctgagca cagcagggga ggggttaatg cagatggcaa gtgcaccaag gagaaggcag
1441 gaatgctgga gcctggaata agggaggaga ggggactgga gagtgtgggg aataggaaga
1501 agaaatttcc tttagactaa cgaatatatt gggggggagga atagagggga ggtgtgcagg
1561 aaccagcaat gagaaggcca ggaaaagaaa gagctgaaaa tgcagaaagc cgaagagtta
1621 gaacttttgg atacagcaga agaaacagcg gctccactac cgacctgccc ccggttcgat
1681 gtccttccaa gaatgaagtc tttccctggt gatggtcccc tgccctgtct ttccagcatc
1741 cactctgtct tgtcctcctg gaagtgtatc tcagtcagcc agtggcttct tgatgatggc
1801 ggtggaggtg gtggttgtag tgtgatggat ccccttttagg ttatttaggg gtatatgtcc
1861 cctgcttgaa ccctgaaggc caggtaatga gccatggcca ttgtccccag ctgaggacca
1921 ggtgtctcta aaaacccaaa catcctggag agtatgcgag aacctaccaa gaaaaacagt
1981 ctcattactc atatacagca ggcaaagaga cagaaaatta actgaaaagc agtttagaga
2041 ctgggggagg ccggatctct agagccatcc tgctgagtgc cctgtgtgta agtcctaata
2101 aactcaccta ctcaccaa
```

APOL3 cDNA SEQUENCE 2 (SEQ ID NO: 3)

NM_014349 Homo sapiens apolipoprotein L, 3 (APOL3), transcript variant alpha/a, mRNA

```
   1 agcaggaggg tgggaccaag ggtgctgctg gaccaaggat gggactgggc caagggtggg
  61 gctgggaagc atcctgtttt gcatgtttga tcaggagctg ctgccaagtt gtgactttca
 121 ctttcccttt tgggttccag ggtatatctc agagcctgga gaacgtgtct ggttattatg
 181 cagatgcacg gctggaggtg ggatccacac agctcagaac agctggatct tgctcacact
```

```
 241 ctttcaagag aagcttcctt ggacaaaagg accctgcctt ggtgtgagag tgagggcaga
 301 gggagctgga gcaagtagaa tttctctaaa taccagctgg ctggggccca ggagattaaa
 361 aaacaccggg ctaggttggt cttggcattt gctgacacgc aaagggattg cagagatcca
 421 gcccctccaa cctccctctg tccacaggtg gctcacattc agtcccacaa tttgctttct
 481 cctcctcaag ggttaagaaa aaaaacgaac ccttccagtc aggtcagtga ctggagagct
 541 ccatggaaag tctctcagtg acctggctgc tggcaccatg gactcagaaa agaaacgctt
 601 tactgaagag gccaccaaat acttccggga gagagtcagc ccagtgcatc tgcaaatcct
 661 gctgactaac aatgaagcct ggaagagatt cgtgactgcg gctgaattgc ccagggatga
 721 ggcagatgct ctctacgaag ctctgaagaa gcttagaaca tatgcagcta ttgaggacga
 781 atatgtgcag cagaaagatg agcagtttag ggaatggttt ttgaaagagt ttccccaagt
 841 caagaggaag atccaggagt ccatagaaaa gcttcgtgcc cttgcaaatg gtattgaaga
 901 ggtccacaga ggctgcacca tctccaatgt ggtgtccagc tccactggcg ctgcctctgg
 961 catcatgtcc cttgctggtc ttgttttggc accatttaca gcagggacga gtctggccct
1021 tactgcagct ggggtagggc tgggagcagc gtctgctgtg actgggatca ccaccagcat
1081 cgtggagcac tcatacacat catcagcaga agctgaagcc agcaggctga ctgcaaccag
1141 cattgaccga ttgaaggtat ttaaggaagt tatgcgtgac atcacaccca acttactttc
1201 ccttcttaat aattattacg aagccacaca aaccattggg agtgaaatcc gtgccatcag
1261 gcaagccaga gccagggccc gactccctgt gaccacctgg cgaatctcag ctggaagtgg
1321 tggtcaagca gagagaacga ttgcaggcac cacccgggca gtgagcagag gagcccggat
1381 cctgagtgcg accacttcag gcatcttcct tgcactggat gtggtcaacc ttgtatacga
1441 gtcaaagcac ttgcatgagg gggcaaagtc tgcatctgct gaggagctga ggcggcaggc
1501 tcaggagctg gaggagaatc taatggagct cactcagatc tatcagcgtc tgaatccatg
1561 ccatacccac tgaccccaga ccagtgcagc cagcagggga ggtgagccat acacaggcca
1621 cgacaaaatg caggcatttt attaggggga taaagagggc aaggtaaagt ttatggagct
1681 gagtgttagt gactttggca tttctgtagc tgagcacagc aggggagggg ttaatgcaga
1741 tggcaagtgc accaaggaga aggcaggaat gctggagcct ggaataaggg aggagagggg
1801 actggagagt gtggggaata ggaagaagaa atttccttta gactaacgaa tatattgggg
1861 ggaggaatag aggggaggtg tgcaggaacc agcaatgaga aggccaggaa aagaaagagc
1921 tgaaaatgca gaaagccgaa gagttagaac ttttggatac agcagaagaa acagcggctc
1981 cactaccgac ctgcccccgg ttcgatgtcc ttccaagaat gaagtctttc cctggtgatg
2041 gtcccctgcc ctgtctttcc agcatccact ctgtcttgtc ctcctggaag tgtatctcag
2101 tcagccagtg gcttcttgat gatggcggtg gaggtggtgg ttgtagtgtg atggatcccc
2161 tttaggttat ttaggggtat atgtcccctg cttgaaccct gaaggccagg taatgagcca
2221 tggccattgt ccccagctga ggaccaggtg tctctaaaaa cccaaacatc ctggagagta
2281 tgcgagaacc taccaagaaa aacagtctca ttactcatat acagcaggca aagagacaga
2341 aaattaactg aaaagcagtt tagagactgg gggaggccgg atctctagag ccatcctgct
2401 gagtgccctg tgtgtaagtc ctaataaact cacctactca ccaa
```

<u>APOL3 cDNA SEQUENCE 3 (SEQ ID NO: 4)</u>

NM_030644 Homo sapiens apolipoprotein L, 3 (APOL3), transcript variant alpha/b, mRNA

```
   1 agcaggaggg tgggaccaag ggtgctgctg gaccaaggat gggactgggc caagggtggg
  61 gctgggaagc atcctgtttt gcatgtttga tcaggagctg ctgccaagtt gtgactttca
 121 ctttcccttt tgggttccag ggtatatctc agagcctgga gaacgtgtct ggttattatg
 181 cagatgcacg gctggaggtg ggatccacac agctcagaac agctggatct tgctcacact
 241 ctttcaagag aagcttcctt ggacaaaagg accctgcctt ggtgtgagag tgagggcaga
 301 gggagctgga gcaagtagaa tttctctaaa taccagctgg ctggggccca ggagattaaa
 361 aaacaccggg ctagggttaa gaaaaaaaac gaacccttcc agtcaggtca gtgactggag
 421 agctccatgg aaagtctctc agtgacctgg ctgctggcac catggactca gaaaagaaac
 481 gctttactga agaggccacc aaatacttcc gggagagagt cagcccagtg catctgcaaa
 541 tcctgctgac taacaatgaa gcctggaaga gattcgtgac tgcggctgaa ttgcccaggg
 601 atgaggcaga tgctctctac gaagctctga agaagcttag aacatatgca gctattgagg
 661 acgaatatgt gcagcagaaa gatgagcagt ttagggaatg gtttttgaaa gagtttcccc
 721 aagtcaagag gaagatccag gagtccatag aaaagcttcg tgcccttgca aatggtattg
 781 aagaggtcca cagaggctgc accatctcca atgtggtgtc cagctccact ggcgctgcct
 841 ctggcatcat gtcccttgct ggtcttgttt tggcaccatt tacagcaggg acgagtctgg
 901 cccttactgc agctggggta gggctgggag cagcgtctgc tgtgactggg atcaccacca
 961 gcatcgtgga gcactcatac acatcatcag cagaagctga agccagcagg ctgactgcaa
1021 ccagcattga ccgattgaag gtatttaagg aagttatgcg tgacatcaca cccaacttac
1081 tttcccttct taataattat tacgaagcca cacaaaccat tgggagtgaa atccgtgcca
1141 tcaggcaagc cagagccagg gcccgactcc ctgtgaccac ctggcgaatc tcagctggaa
```

```
1201 gtggtggtca agcagagaga acgattgcag gcaccacccg ggcagtgagc agaggagccc
1261 ggatcctgag tgcgaccact tcaggcatct tccttgcact ggatgtggtc aaccttgtat
1321 acgagtcaaa gcacttgcat gagggggcaa agtctgcatc tgctgaggag ctgaggcggc
1381 aggctcagga gctggaggag aatctaatgg agctcactca gatctatcag cgtctgaatc
1441 catgccatac ccactgaccc cagaccagtg cagccagcag ggaggtgag ccatacacag
1501 gccacgacaa aatgcaggca ttttattagg gggataaaga gggcaaggta aagtttatgg
1561 agctgagtgt tagtgacttt ggcatttctg tagctgagca cagcagggga ggggttaatg
1621 cagatggcaa gtgcaccaag gagaaggcag gaatgctgga gcctggaata agggaggaga
1681 ggggactgga gagtgtgggg aataggaaga agaaatttcc tttagactaa cgaatatatt
1741 gggggggagga atagagggga ggtgtgcagg aaccagcaat gagaaggcca ggaaaagaaa
1801 gagctgaaaa tgcagaaagc cgaagagtta gaacttttgg atacagcaga agaaacagcg
1861 gctccactac cgacctgccc ccggttcgat gtccttccaa gaatgaagtc tttccctggt
1921 gatggtcccc tgccctgtct ttccagcatc cactctgtct tgtcctcctg gaagtgtatc
1981 tcagtcagcc agtggcttct tgatgatggc ggtggaggtg gtggttgtag tgtgatggat
2041 cccctttagg ttatttaggg gtatatgtcc cctgcttgaa ccctgaaggc caggtaatga
2101 gccatggcca ttgtccccag ctgaggacca ggtgtctcta aaaacccaaa catcctggag
2161 agtatgcgag aacctaccaa gaaaaacagt ctcattactc atatacagca ggcaaagaga
2221 cagaaaatta actgaaaagc agtttagaga ctgggggagg ccggatctct agagccatcc
2281 tgctgagtgc cctgtgtgta agtcctaata aactcaccta ctcaccaa
```

<u>APOL3 cDNA SEQUENCE 4 (SEQ ID NO: 5)</u>

NM_030644 Homo sapiens apolipoprotein L, 3 (APOL3), transcript variant alpha/c, mRNA

```
   1 agcaggaggg tgggaccaag ggtgctgctg gaccaaggat gggactgggc caagggtggg
  61 gctgggaagc atcctgtttt gcatgtttga tcaggagctg ctgccaagtt gtgactttca
 121 ctttcccttt tgggttccag ggtatatctc agagcctgga gaacgtgtct ggttattatg
 181 cagatgcacg gctggaggtg ggatccacac agctcagaac agctggatct tgctcacact
 241 ctttcaagag aagcttcctt gggttaagaa aaaaaacgaa cccttccagt caggtcagtg
 301 actggagagc tccatggaaa gtctctcagt gacctggctg ctggcaccat ggactcagaa
 361 aagaaacgct ttactgaaga ggccaccaaa tacttccggg agagagtcag cccagtgcat
 421 ctgcaaatcc tgctgactaa caatgaagcc tggaagagat tcgtgactgc ggctgaattg
 481 cccagggatg aggcagatgc tctctacgaa gctctgaaga agcttagaac atatgcagct
 541 attgaggacg aatatgtgca gcagaaagat gagcagttta gggaatggtt tttgaaagag
 601 tttccccaag tcaagaggaa gatccaggag tccatagaaa agcttcgtgc ccttgcaaat
 661 ggtattgaag aggtccacag aggctgcacc atctccaatg tggtgtccag ctccactggc
 721 gctgcctctg gcatcatgtc ccttgctggt cttgttttgg caccatttac agcagggacg
 781 agtctggccc ttactgcagc tggggtaggg ctgggagcag cgtctgctgt gactgggatc
 841 accaccagca tcgtggagca ctcatacaca tcatcagcag aagctgaagc cagcaggctg
 901 actgcaacca gcattgaccg attgaaggta tttaaggaag ttatgcgtga catcacaccc
 961 aacttacttt cccttcttaa taattattac gaagccacac aaaccattgg gagtgaaatc
1021 cgtgccatca ggcaagccag agccagggcc cgactccctg tgaccacctg gcgaatctca
1081 gctggaagtg gtggtcaagc agagagaacg attgcaggca ccacccgggc agtgagcaga
1141 ggagcccgga tcctgagtgc gaccacttca ggcatcttcc ttgcactgga tgtggtcaac
1201 cttgtatacg agtcaaagca cttgcatgag gggcaaagt ctgcatctgc tgaggagctg
1261 aggcggcagg ctcaggagct ggaggagaat ctaatggagc tcactcagat ctatcagcgt
1321 ctgaatccat gccataccca ctgaccccag accagtgcag ccagcagggg aggtgagcca
1381 tacacaggcc acgacaaaat gcaggcattt tattaggggg ataaagaggg caaggtaaag
1441 tttatggagc tgagtgttag tgactttggc atttctgtag ctgagcacag caggggaggg
1501 gttaatgcag atggcaagtg caccaaggag aaggcaggaa tgctggagcc tggaataagg
1561 gaggagaggg gactggagag tgtggggaat aggaagaaga aatttccttt agactaacga
1621 atatattggg gggaggaata gaggggaggt gtgcaggaac cagcaatgag aaggccagga
1681 aaagaaagag ctgaaaatgc agaaagccga agagttagaa cttttggata cagcagaaga
1741 aacagcggct ccactaccga cctgcccccg gttcgatgtc cttccaagaa tgaagtcttt
1801 ccctggtgat ggtcccctgc cctgtctttc cagcatccac tctgtcttgt cctcctggaa
1861 gtgtatctca gtcagccagt ggcttcttga tgatggcggt ggaggtggtg gttgtagtgt
1921 gatggatccc ctttaggtta tttaggggta tatgtcccct gcttgaaccc tgaaggccag
1981 gtaatgagcc atggccattg tccccagctg aggaccaggt gtctctaaaa acccaaacat
2041 cctggagagt atgcgagaac ctaccaagaa aaacagtctc attactcata tacagcaggc
2101 aaagagacag aaaattaact gaaaagcagt ttagagactg gggaggccg gatctctaga
2161 gccatcctgc tgagtgccct gtgtgtaagt cctaataaac tcacctactc accaa
```

APOL3 cDNA SEQUENCE 5 (SEQ ID NO: 6)

NM_145641 Homo sapiens apolipoprotein L, 3 (APOL3), transcript variant beta/a, mRNA

```
   1 actctgggga aaggagggta caaccacatg taaattttat tataatgatg gtataatgaa
  61 cttggggtga cctgagggca tgtttttgtg ttactgggca tgtgcccctt taggagactt
 121 ccacctgtgc cttactttct ctttctttgg gatgtgctgg ccacagactt taccaaaaac
 181 tccatccatt aggatggcat taggacggtt ctggggctaa acttaggtgg gccaggggct
 241 gtttcactgt cagcctttct actctctttt cttacccact cctagctgct aatgtctatt
 301 taactaccta atatttcccc ctttagagaa aaaaagccaa atttttgggt agatcggtga
 361 caattaatct ggctacttcc tgctgacaag aggcagtggt aataattggg ttctcttttt
 421 tgctctcttg tagctggtag gttgggcaga gaaaagtggt ggccatccaa ggggcccacg
 481 tagatatcag acatggttga gacctcgcgg taaccttgtg tagaatcatt tggagtttta
 541 tggattctag gtgggaagaa acaaaacaac cttgtaaatc aaatgagcat cgtttgaaag
 601 ctataagttg tataaagctg ttttaggacc aagaaagggg gctaaccagg aaaaccagga
 661 ccagttgtta aatttccacc agtcaaagcc tcctgaaact ctgttttcca ttaacttgtt
 721 ggccctgtct gtaatttttt taagttggtt tgcactttac ctgattggtt gatgaaaaca
 781 gcaatgttta tcaagtgttg cacaagctcc cccttgattg gctgtgagca aattaaaagc
 841 tcatcaattt cataagacta tgcttgctaa tgaagcaatt tgttctgaaa gggtattgac
 901 ctggttagtt agattaataa tttgttgagt aatttttaag aagagtttct gggcagtaaa
 961 aatggagtta aggaggtcct ccagttcccg tgccaatacc agccagaacg ccaattatag
1021 tcagtgctgt taaaactcct ctttcagttt taccagattg agcgtgtata gggaggggaa
1081 tgctatcgat aagagtgagt ttggggatga tgtaaactag ggccaaatcc ccgtttaatt
1141 agcagacaag cagagatatg ccttgtttct atacaaaaat gtgatttgtc atgttaagac
1201 aaatatcaac agtgatacta aagtagggtt tttccatgct gtgtaagtct gctttaatac
1261 cctcagaagt ggaagtcaag gctagttcat gtaatggggg aacacaggca gtacttgaca
1321 gaatgaaaga ggaatcaaaa gcccaattag aggaaggagt gatgggatcc caaccaatat
1381 tgtgaaatta tggggagctg cagttttttg caatgatttt gcccgaagtt gtgagactga
1441 aaccagaagt tgttatgttt aaaatgatat tctggtgtgg gtctggggaa aagggtagat
1501 ccagaatagc tggcttcttt ccacatagtg gttctgttgg ggggcttggg aaatgaataa
1561 aacacaaaga agaattagaa tatcaggtga aggtagcagg tgctcctggc agaagaaccc
1621 acataaggaa gtgtccagaa gccacacagg gtatatctca gagcctggag aacgtgtctg
1681 gttattatgc agatgcacgg ctggaggtgg gatccacaca gctcagaaca gctggatctt
1741 gctcacactc tttcaagaga agcttccttg aaaagaaacg ctttactgaa gaggccacca
1801 aatacttccg ggagagagtc agcccagtgc atctgcaaat cctgctgact aacaatgaag
1861 cctggaagag attcgtgact gcggctgaat tgcccaggga tgaggcagat gctctctacg
1921 aagctctgaa gaagcttaga acatatgcag ctattgagga cgaatatgtg cagcagaaag
1981 atgagcagtt tagggaatgg tttttgaaag agtttcccca agtcaagagg aagatccagg
2041 agtccataga aaagcttcgt gcccttgcaa atggtattga agaggtccac agaggctgca
2101 ccatctccaa tgtggtgtcc agctccactg gcgctgcctc tggcatcatg tcccttgctg
2161 gtcttgtttt ggcaccattt acagcaggga cgagtctggc ccttactgca gctggggtag
2221 ggctgggagc agcgtctgct gtgactggga tcaccaccag catcgtggag cactcataca
2281 catcatcagc agaagctgaa gccagcaggc tgactgcaac cagcattgac cgattgaagg
2341 tatttaagga agttatgcgt gacatcacac ccaacttact ttcccttctt aataattatt
2401 acgaagccac acaaaccatt gggagtgaaa tccgtgccat caggcaagcc agagccaggg
2461 cccgactccc tgtgaccacc tggcgaatct cagctggaag tggtggtcaa gcagagagaa
2521 cgattgcagg caccacccgg gcagtgagca gaggagcccg gatcctgagt gcgaccactt
2581 caggcatctt ccttgcactg gatgtggtca accttgtata cgagtcaaag cacttgcatg
2641 aggggggcaaa gtctgcatct gctgaggagc tgaggcggca ggctcaggag ctggaggaga
2701 atctaatgga gctcactcag atctatcagc gtctgaatcc atgccatacc cactgacccc
2761 agaccagtgc agccagcagg ggaggtgagc catacacagg ccacgacaaa atgcaggcat
2821 tttattaggg ggataaagag ggcaaggtaa agtttatgga gctgagtgtt agtgactttg
2881 gcatttctgt agctgagcac agcaggggag gggttaatgc agatggcaag tgcaccaagg
2941 agaaggcagg aatgctggag cctggaataa gggaggagag gggactggag agtgtgggga
3001 ataggaagaa gaaatttcct ttagactaac gaatatattg gggggaggaa tagaggggag
3061 gtgtgcagga accagcaatg agaaggccag gaaaagaaag agctgaaaat gcagaaagcc
3121 gaagagttag aacttttgga tacagcagaa gaaacagcgg ctccactacc gacctgcccc
3181 cggttcgatg tccttccaag aatgaagtct ttccctggtg atggtcccct gccctgtctt
3241 tccagcatcc actctgtctt gtcctcctgg aagtgtatct cagtcagcca gtggcttctt
3301 gatgatggcg gtggaggtgg tggttgtagt gtgatggatc ccctttaggt tatttagggg
3361 tatatgtccc ctgcttgaac cctgaaggcc aggtaatgag ccatggccat tgtccccagc
3421 tgaggaccag gtgtctctaa aaacccaaac atcctggaga gtatgcgaga acctaccaag
3481 aaaaacagtc tcattactca tatacagcag gcaaagagac agaaaattaa ctgaaaagca
```

```
3541 gtttagagac tgggggaggc cggatctcta gagccatcct gctgagtgcc ctgtgtgtaa
3601 gtcctaataa actcacctac tcaccaa
```

APOL3 cDNA SEQUENCE 6 (SEQ ID NO: 7)

NM_145642 Homo sapiens apolipoprotein L, 3 (APOL3), transcript variant beta/a, mRNA

```
   1 actctgggga aaggagggta caaccacatg taaattttat tataatgatg gtataatgaa
  61 cttggggtga cctgagggca tgtttttgtg ttactgggca tgtgcccctt taggagactt
 121 ccacctgtgc cttactttct ctttctttgg gatgtgctgg ccacagactt taccaaaaac
 181 tccatccatt aggatggcat taggacggtt ctggggctaa acttaggtgg gccaggggct
 241 gtttcactgt cagcctttct actctctttt cttaccccact cctagctgct aatgtctatt
 301 taactaccta atatttcccc ctttagagaa aaaaagccaa atttttgggt agatcggtga
 361 caattaatct ggctacttcc tgctgacaag aggcagtggt aataattggg ttctcttttt
 421 tgctctcttg tagctggtag gttgggcaga gaaaagtggt ggccatccaa ggggcccacg
 481 tagatatcag acatggttga gacctcgcgg taaccttgtg tagaatcatt tggagtttta
 541 tggattctag gtgggaagaa acaaaacaac cttgtaaatc aaatgagcat cgtttgaaag
 601 ctataagttg tataaagctg ttttaggacc aagaaagggg gctaaccagg aaaaccagga
 661 ccagttgtta aatttccacc agtcaaagcc tcctgaaact ctgttttcca ttaacttgtt
 721 ggccctgtct gtaatttttt taagttggtt tgcactttac ctgattggtt gatgaaaaca
 781 gcaatgttta tcaagtgttg cacaagctcc cccttgattg gctgtgagca aattaaaagc
 841 tcatcaattt cataagacta tgcttgctaa tgaagcaatt tgttctgaaa gggtattgac
 901 ctggttagtt agattaataa tttgttgagt aatttttaag aagagtttct gggcagtaaa
 961 aatggagtta aggaggtcct ccagttcccg tgccaatacc agccagaacg ccaattatag
1021 tcagtgctgt taaaactcct ctttcagttt taccagattg agcgtgtata gggaggggaa
1081 tgctatcgat aagagtgagt ttggggatga tgtaaactag ggccaaatcc ccgtttaatt
1141 agcagacaag cagagatatg ccttgtttct atacaaaaat gtgatttgtc atgttaagac
1201 aaatatcaac agtgatacta aagtagggtt tttccatgct gtgtaagtct gctttaatac
1261 cctcagaagt ggaagtcaag gctagttcat gtaatggggg aacacaggca gtacttgaca
1321 gaatgaaaga ggaatcaaaa gcccaattag aggaaggagt gatgggatcc caaccaatat
1381 tgtgaaatta tggggagctg cagttttttg caatgatttt gcccgaagtt gtgagactga
1441 aaccagaagt tgttatgttt aaaatgatat tctggtgtgg gtctggggaa aagggtagat
1501 ccagaatagc tggcttcttt ccacatagtg gttctgttgg ggggcttggg aaatgaataa
1561 aacacaaaga agaattagaa tatcagggta tatctcagag cctggagaac gtgtctggtt
1621 attatgcaga tgcacggctg gaggtgggat ccacacagct cagaacagct ggatcttgct
1681 cacactcttt caagagaagc ttccttgaaa agaaacgctt tactgaagag gccaccaaat
1741 acttccggga gagagtcagc ccagtgcatc tgcaaatcct gctgactaac aatgaagcct
1801 ggaagagatt cgtgactgcg gctgaattgc ccagggatga ggcagatgct ctctacgaag
1861 ctctgaagaa gcttagaaca tatgcagcta ttgaggacga atatgtgcag cagaaagatg
1921 agcagtttag ggaatggttt ttgaaagagt ttccccaagt caagaggaag atccaggagt
1981 ccatagaaaa gcttcgtgcc cttgcaaatg gtattgaaga ggtccacaga ggctgcacca
2041 tctccaatgt ggtgtccagc tccactggcg ctgcctctgg catcatgtcc cttgctggtc
2101 ttgttttggc accatttaca gcagggacga gtctggccct tactgcagct ggggtagggc
2161 tgggagcagc gtctgctgtg actgggatca ccaccagcat cgtggagcac tcatacacat
2221 catcagcaga agctgaagcc agcaggctga ctgcaaccag cattgaccga ttgaaggtat
2281 ttaaggaagt tatgcgtgac atcacaccca acttactttc ccttcttaat aattattacg
2341 aagccacaca aaccattggg agtgaaatcc gtgccatcag gcaagccaga gccagggccc
2401 gactccctgt gaccacctgg cgaatctcag ctggaagtgg tggtcaagca gagagaacga
2461 ttgcaggcac cacccgggca gtgagcagag gagcccggat cctgagtgcg accacttcag
2521 gcatcttcct tgcactggat gtggtcaacc ttgtatacga gtcaaagcac ttgcatgagg
2581 gggcaaagtc tgcatctgct gaggagctga ggcggcaggc tcaggagctg gaggagaatc
2641 taatggagct cactcagatc tatcagcgtc tgaatccatg ccatacccac tgaccccaga
2701 ccagtgcagc cagcagggga ggtgagccat acacaggcca cgacaaaatg caggcatttt
2761 attaggggga taaagagggc aaggtaaagt ttatggagct gagtgttagt gactttggca
2821 tttctgtagc tgagcacagc aggggagggg ttaatgcaga tggcaagtgc accaaggaga
2881 aggcaggaat gctggagcct ggaataaggg aggagagggg actggagagt gtggggaata
2941 ggaagaagaa atttccttta gactaacgaa tatattgggg ggaggaatag aggggaggtg
3001 tgcaggaacc agcaatgaga aggccaggaa aagaaagagc tgaaaatgca gaaagccgaa
3061 gagttagaac ttttggatac agcagaagaa acagcggctc cactaccgac ctgcccccgg
3121 ttcgatgtcc ttccaagaat gaagtctttc cctggtgatg gtcccctgcc ctgtctttcc
3181 agcatccact ctgtcttgtc ctcctggaag tgtatctcag tcagccagtg gcttcttgat
3241 gatggcggtg gaggtggtgg ttgtagtgtg atggatcccc tttaggttat ttaggggtat
```

```
3301 atgtcccctg cttgaaccct gaaggccagg taatgagcca tggccattgt ccccagctga
3361 ggaccaggtg tctctaaaaa cccaaacatc ctggagagta tgcgagaacc taccaagaaa
3421 aacagtctca ttactcatat acagcaggca aagagacaga aaattaactg aaaagcagtt
3481 tagagactgg gggaggccgg atctctagag ccatcctgct gagtgccctg tgtgtaagtc
3541 ctaataaact cacctactca ccaa
```

[0243] The following are human polypeptide sequences for isoform 1, isoform 2 and isoform 3 of APOL3.

APOL3 AMINO ACID SEQUENCE 1 (SEQ ID NO: 8)

NP_663615 Homo sapiens apolipoprotein L, 3 (APOL3), isoform 1, protein

MGLGQGWGWEASCFACLIRSCCQVVTFTFPFGFQGISQSLENVSGYYADARLEVGSTQ
LRTAGSCSHSFKRSFLEKKRFTEEATKYFRERVSPVHLQILLTNNEAWKRFVTAAELPR
DEADALYEALKKLRTYAAIEDEYVQQKDEQFREWFLKEFPQVKRKIQESIEKLRALAN
GIEEVHRGCTISNVVSSSTGAASGIMSLAGLVLAPFTAGTSLALTAAGVGLGAASAVTGI
TTSIVEHSYTSSAEAEASRLTATSIDRLKVFKEVMRDITPNLLSLLNNYYEATQTIGSEIR
AIRQARARARLPVTTWRISAGSGGQAERTIAGTTRAVSRGARILSATTSGIFLALDVVNL
VYESKHLHEGAKSASAEELRRQAQELEENLMELTQIYQRLNPCHTH

APOL3 AMINO ACID SEQUENCE 2 (SEQ ID NO: 9)

NP_055164 Homo sapiens apolipoprotein L, 3 (APOL3), isoform 2, protein

MDSEKKRFTEEATKYFRERVSPVHLQILLTNNEAWKRFVTAAELPRDEADALYEALKK
LRTYAAIEDEYVQQKDEQFREWFLKEFPQVKRKIQESIEKLRALANGIEEVHRGCTISNV
VSSSTGAASGIMSLAGLVLAPFTAGTSLALTAAGVGLGAASAVTGITTSIVEHSYTSSAE
AEASRLTATSIDRLKVFKEVMRDITPNLLSLLNNYYEATQTIGSEIRAIRQARARARLPVT
TWRISAGSGGQAERTIAGTTRAVSRGARILSATTSGIFLALDVVNLVYESKHLHEGAKS
ASAEELRRQAQELEENLMELTQIYQRLNPCHTH

APOL3 AMINO ACID SEQUENCE 3 (SEQ ID NO: 10)

NP_663616 Homo sapiens apolipoprotein L, 3 (APOL3), isoform 3, protein

MSLAGLVLAPFTAGTSLALTAAGVGLGAASAVTGITTSIVEHSYTSSAEAEASRLTATSI
DRLKVFKEVMRDITPNLLSLLNNYYEATQTIGSEIRAIRQARARARLPVTTWRISAGSGG
QAERTIAGTTRAVSRGARILSATTSGIFLALDVVNLVYESKHLHEGAKSASAEELRRQA
QELEENLMELTQIYQRLNPCHTH

[0244] Modifications may be made to the foregoing without departing from the basic aspects of the invention. Although the invention has been described in substantial detail with reference to one or more specific embodiments, those of skill in the art will recognize that changes may be made to the embodiments specifically disclosed in this application, yet these modifications and improvements are

within the scope and spirit of the invention, as set forth in the aspects which follow. All publications or patent documents cited in this specification are incorporated herein by reference as if each such publication or document was specifically and individually indicated to be incorporated herein by reference.

[0245] Citation of the above publications or documents is not intended as an admission that any of the foregoing is pertinent prior art, nor does it constitute any admission as to the contents or date of these publications or documents. U.S. patents and other publications referenced herein are hereby incorporated by reference.

What is claimed is:

1. A method for identifying a subject at risk of osteoarthritis, which comprises detecting the presence or absence of one or more polymorphic variations associated with osteoarthritis in a nucleic acid sample from a subject, wherein the one or more polymorphic variations are detected in a nucleotide sequence selected from the group consisting of:

(a) a nucleotide sequence in SEQ ID NO: 1-7;

(b) a nucleotide sequence which encodes a polypeptide encoded by a nucleotide sequence in SEQ ID NO: 1-7;

(c) a nucleotide sequence which encodes a polypeptide that is 90% or more identical to the amino acid sequence encoded by a nucleotide sequence in SEQ ID NO: 1-7;

(d) a fragment of a nucleotide sequence of (a), (b), or (c);

whereby the presence of the polymorphic variation is indicative of the subject being at risk of osteoarthritis.

2. The method of claim 1, which further comprises obtaining the nucleic acid sample from the subject.

3. The method of claim 1, wherein the one or more polymorphic variations are detected within a region spanning chromosome positions 34828750 and 34833750 in human genomic DNA.

4. The method of claim 1, wherein the one or more polymorphic variations are detected at one or more positions selected from the group consisting of 201, 425, 1095, 2201, 7879, 8395, 8461, 9503, 10304, 10695, 16300, 16444, 17591, 17988, 19116, 19358, 20300, 20669, 20891, 21451, 21978, 22785, 24248, 24770, 24844, 25066, 25096, 25309, 25344, 25529, 25537, 25554, 27963, 28134, 28356, 29648, 29986, 30217, 30267, 30315, 30585, 30724, 30897, 30931, 31080, 31246, 31373, 31463, 31467, 32188, 32288, 32520, 32594, 32657, 32677, 32764, 32784, 32830, 32872, 33121, 33348, 33952, 34184, 34361, 35026, 35192, 35600, 36033, 36289, 38869, 39629, 40530, 41621, 42379, 42802, 42865, 43644, 45051, 45828, 45829, 46257, 47286, 47427, 47963, 48013, 48229, 48282, 48376, 48404, 49900, 52699, 52897, 53414, 53487, 54112, 55492, 59766, 60307, 60701, 60952, 61401, 62379, 62870, 62879, 63499, 64284, 64408, 64760, 65230, 66127, , 6634, 66686, 66694, 67113, 67257, 67403, 67609, 68418, 68610, 69629, 70024, 70848, 71428, 71553, 71633, 71768 ,71769, 73039, 73325, 73412, 73547, 73769, 73806, 74467, 74472, 74473, 74482, 74494, 74592, 74670, 74672, 74714, 74723, 74749, 74861, 74892, 74893, 75176, 75705, 75989, 76027, 77949, 77974, 78167, 78310, 78415, 78575, 78590, 78709, 78875, 79864, 81316, 81320, 81409, 81737, 81843, 82102, 82833, 83461, 83624, 83660, 83701, 83708, 83782, 85707, 85717,

86486, 86833, 87115, 87234, 87479, 87561, 87604, 87674, 87958, 87992, 88019, 88074, 88079, 88115, 88118, 88120, 88135, 88142, 88143, 88149, 88340, 88344, 88512, 88521, 88650, 88827, 89230, 89236, 90754, 90984, 91110, 92026, 92954, 93375, 93794, 94937, 95068, 96188, 97092 and 98812.

5. The method of claim 1, wherein the one or more polymorphic variations are detected at one or more positions in SEQ ID NO: 1 selected from the group consisting of 20300, 87958, 89236, 30267, 32657, 36289, 38869, 45051, 46257, 54112, 60307, 63499, 20891, 52699 and 71768.

6. The method of claim 1, wherein the one or more polymorphic variations are detected at one or more positions in linkage disequilibrium with one or more positions in claim 3, 4 or 5.

7. The method of claim 1, wherein detecting the presence or absence of the one or more polymorphic variations comprises:

hybridizing an oligonucleotide to the nucleic acid sample, wherein the oligonucleotide is complementary to a nucleotide sequence in the nucleic acid and hybridizes to a region adjacent to the polymorphic variation;

extending the oligonucleotide in the presence of one or more nucleotides, yielding extension products; and

detecting the presence or absence of a polymorphic variation in the extension products.

8. The method of claim 1, wherein the subject is a human.

9. The method of claim 8, wherein the subject is a human female.

10. The method of claim 8, wherein the subject is a human male.

11. A method for identifying a polymorphic variation associated with osteoarthritis proximal to an incident polymorphic variation associated with osteoarthritis, which comprises:

identifying a polymorphic variation proximal to the incident polymorphic variation associated with osteoarthritis, wherein the polymorphic variation is detected in a nucleotide sequence selected from the group consisting of:

(a) a nucleotide sequence in SEQ ID NO: 1-7;

(b) a nucleotide sequence which encodes a polypeptide encoded by a nucleotide sequence in SEQ ID NO: 1-7;

(c) a nucleotide sequence which encodes a polypeptide that is 90% or more identical to the amino acid sequence encoded by a nucleotide sequence in SEQ ID NO: 1-7;

(d) a fragment of a nucleotide sequence of (a), (b), or (c) comprising a polymorphic variation;

determining the presence or absence of an association of the proximal polymorphic variant with osteoarthritis.

12. The method of claim 11, wherein the incident polymorphic variation is at one or more positions in claim 3, 4 or 5.

13. The method of claim 11, wherein the proximal polymorphic variation is within a region between about 5 kb 5' of the incident polymorphic variation and about 5 kb 3' of the incident polymorphic variation.

14. The method of claim 11, which further comprises determining whether the proximal polymorphic variation is in linkage disequilibrium with the incident polymorphic variation.

15. The method of claim 11, which further comprises identifying a second polymorphic variation proximal to the identified proximal polymorphic variation associated with osteoarthritis and determining if the second proximal polymorphic variation is associated with osteoarthritis.

16. The method of claim 15, wherein the second proximal polymorphic variant is within a region between about 5 kb 5' of the incident polymorphic variation and about 5 kb 3' of the proximal polymorphic variation associated with osteoarthritis.

17. An isolated nucleic acid comprising a nucleotide sequence selected from the group consisting of:

(a) a nucleotide sequence in SEQ ID NO: 1-7;

(b) a nucleotide sequence which encodes a polypeptide encoded by a nucleotide sequence in SEQ ID NO: 1-7;

(c) a nucleotide sequence which encodes a polypeptide that is 90% or more identical to the amino acid sequence encoded by a nucleotide sequence in SEQ ID NO: 1-7;

(d) a fragment of a nucleotide sequence of (a), (b), or (c) comprising a polymorphic variation; and

(e) a nucleotide sequence complementary to the nucleotide sequences of (a), (b), (c), or (d);

wherein the nucleotide sequence comprises a polymorphic variation associated with osteoarthritis selected from the group consisting of an adenine at position 20300, a thymine at position 46257, an adenine at position 89236, a guanine at position 30267, an adenine at position 32657, a cytosine at position 36289, a guanine at position 38869, a thymine at position 45051, a guanine at position 54112, an adenine at position 60307, a thymine at position 63499, a guanine at position 20891, a guanine at position 52699, and a cytosine at position 71768.

18. An oligonucleotide comprising a nucleotide sequence complementary to a portion of the nucleotide sequence of (a), (b), (c), or (d) in claim 17, wherein the 3' end of the oligonucleotide is adjacent to a polymorphic variation associated with osteoarthritis.

19. A microarray comprising an isolated nucleic acid of claim 17 linked to a solid support.

20. An isolated polypeptide encoded by the isolated nucleic acid sequence of claim 17.

21. A method for identifying a candidate therapeutic for treating osteoarthritis, which comprises:

(a) introducing a test molecule to a system which comprises a nucleic acid comprising a nucleotide sequence selected from the group consisting of:

(i) a nucleotide sequence in SEQ ID NO: 1-7;

(ii) a nucleotide sequence which encodes a polypeptide encoded by a nucleotide sequence in SEQ ID NO: 1-7;

(iii) a nucleotide sequence which encodes a polypeptide that is 90% or more identical to the amino acid sequence encoded by a nucleotide sequence in SEQ ID NO: 1-7;

(iv) a fragment of a nucleotide sequence of (a), (b), or (c); or

introducing a test molecule to a system which comprises a protein encoded by a nucleotide sequence of (i), (ii), (iii), or (iv); and

(b) determining the presence or absence of an interaction between the test molecule and the nucleic acid or protein,

whereby the presence of an interaction between the test molecule and the nucleic acid or protein identifies the test molecule as a candidate therapeutic for treating osteoarthritis.

22. The method of claim 21, wherein the system is an animal.

23. The method of claim 21, wherein the system is a cell.

24. The method of claim 21, wherein the nucleotide sequence comprises one or more polymorphic variations associated with osteoarthritis.

25. The method of claim 24, wherein the one or more polymorphic variations associated with osteoarthritis are at one or more positions in claim 3, 4 or 5.

26. A method for treating osteoarthritis in a subject, which comprises contacting one or more cells of a subject in need thereof with a nucleic acid, wherein the nucleic acid comprises a nucleotide sequence selected from the group consisting of:

(a) a nucleotide sequence in SEQ ID NO: 1-7;

(b) a nucleotide sequence which encodes a polypeptide encoded by a nucleotide sequence in SEQ ID NO: 1-7;

(c) a nucleotide sequence which encodes a polypeptide that is 90% or more identical to the amino acid sequence encoded by a nucleotide sequence in SEQ ID NO: 1-7;

(d) a fragment of a nucleotide sequence of (a), (b), or (c); and

(e) a nucleotide sequence complementary to the nucleotide sequences of (a), (b), (c), or (d);

whereby contacting the one or more cells of the subject with the nucleic acid treats the osteoarthritis in the subject.

27. The method of claim 25, wherein the nucleic acid is RNA or PNA.

28. The method of claim 27, wherein the nucleic acid is duplex RNA.

29. A method for treating osteoarthritis in a subject, which comprises contacting one or more cells of a subject in need thereof with a protein, wherein the protein is encoded by a nucleotide sequence which comprises a polynucleotide sequence selected from the group consisting of:

(a) a nucleotide sequence in SEQ ID NO: 1-7;

(b) a nucleotide sequence which encodes a polypeptide encoded by a nucleotide sequence in SEQ ID NO: 1-7;

(c) a nucleotide sequence which encodes a polypeptide that is 90% or more identical to the amino acid sequence encoded by a nucleotide sequence in SEQ ID NO: 1-7;

(d) a fragment of a nucleotide sequence of (a), (b), or (c);

whereby contacting the one or more cells of the subject with the protein treats the osteoarthritis in the subject.

30. A method for treating osteoarthritis in a subject, which comprises:

detecting the presence or absence of one or more polymorphic variations associated with osteoarthritis in a nucleic acid sample from a subject, wherein the one or more polymorphic variation are detected in a nucleotide sequence selected from the group consisting of:

(a) a nucleotide sequence in SEQ ID NO: 1-7;

(b) a nucleotide sequence which encodes a polypeptide encoded by a nucleotide sequence in SEQ ID NO: 1-7;

(c) a nucleotide sequence which encodes a polypeptide that is 90% or more identical to the amino acid sequence encoded by a nucleotide sequence in SEQ ID NO: 1-7;

(d) a fragment of a nucleotide sequence of (a), (b), or (c) comprising a polymorphic variation; and

administering an osteoarthritis treatment to a subject in need thereof based upon the presence or absence of the one or more polymorphic variations in the nucleic acid sample.

31. The method of claim 30, wherein the one or more polymorphic variations are detected at one or more positions in claim 3, 4 or 5.

32. The method of claim 30, wherein the treatment is selected from the group consisting of administering a corticosteroid, a nonsteroidal anti-inflammatory drug (NSAID), a cyclooxygenase-2 (COX-2) inhibitor, an antibody, a glucocorticoid, hyaluronic acid, chondrotin sulfate, glucosamine or acetaminophen; prescribing a heat/cold regimen or a joint protection regimen; performing joint surgery; prescribing a weight control regimen; and combinations of the foregoing.

33. A method for detecting or preventing osteoarthritis in a subject, which comprises:

detecting the presence or absence of one or more polymorphic variations associated with osteoarthritis in a nucleic acid sample from a subject, wherein the polymorphic variation is detected in a nucleotide sequence selected from the group consisting of:

(a) a nucleotide sequence in SEQ ID NO: 1-7;

(b) a nucleotide sequence which encodes a polypeptide encoded by a nucleotide sequence in SEQ ID NO: 1-7;

(c) a nucleotide sequence which encodes a polypeptide that is 90% or more identical to the amino acid sequence encoded by a nucleotide sequence in SEQ ID NO: 1-7;

(d) a fragment of a nucleotide sequence of (a), (b), or (c) comprising a polymorphic variation; and

administering an osteoarthritis prevention or detection procedure to a subject in need thereof based upon the presence or absence of the one or more polymorphic variations in the nucleic acid sample.

34. The method of claim 33, wherein the one or more polymorphic variations are detected at one or more positions in claim 3, 4 or 5.

35. The method of claim 33, wherein the osteoarthritis prevention is selected from the group consisting of administering a corticosteroid, a nonsteroidal anti-inflammatory drug (NSAID), a cyclooxygenase-2 (COX-2) inhibitor, an antibody, a glucocorticoid, hyaluronic acid, chondrotin sulfate, glucosamine or acetaminophen; prescribing a heat/cold regimen or a joint protection regimen; performing joint surgery; prescribing a weight control regimen; and combinations of the foregoing.

36. A method of targeting information for preventing or treating osteoarthritis to a subject in need thereof, which comprises:

detecting the presence or absence of one or more polymorphic variations associated with osteoarthritis in a nucleic acid sample from a subject, wherein the polymorphic variation is detected in a nucleotide sequence selected from the group consisting of:

(a) a nucleotide sequence in SEQ ID NO: 1-7;

(b) a nucleotide sequence which encodes a polypeptide encoded by a nucleotide sequence in SEQ ID NO: 1-7;

(c) a nucleotide sequence which encodes a polypeptide that is 90% or more identical to the amino acid sequence encoded by a nucleotide sequence in SEQ ID NO: 1-7;

(d) a fragment of a nucleotide sequence of (a), (b), or (c) comprising a polymorphic variation; and

directing information for preventing or treating osteoarthritis to a subject in need thereof based upon the presence or absence of the one or more polymorphic variations in the nucleic acid sample.

37. The method of claim 36, wherein the one or more polymorphic variations are detected at one or more positions in claim 3, 4 or 5.

38. A composition comprising a cell from a subject having osteoarthritis or at risk of osteoarthritis and an antibody that specifically binds to a protein, polypeptide or peptide encoded by a nucleotide sequence identical to or 90% or more identical to a nucleotide sequence in SEQ ID NO: 1-7.

39. A composition comprising a cell from a subject having osteoarthritis or at risk of osteoarthritis and a RNA, DNA, PNA or ribozyme molecule comprising a nucleotide sequence identical to or 90% or more identical to a portion of a nucleotide sequence in SEQ ID NO: 1-7.

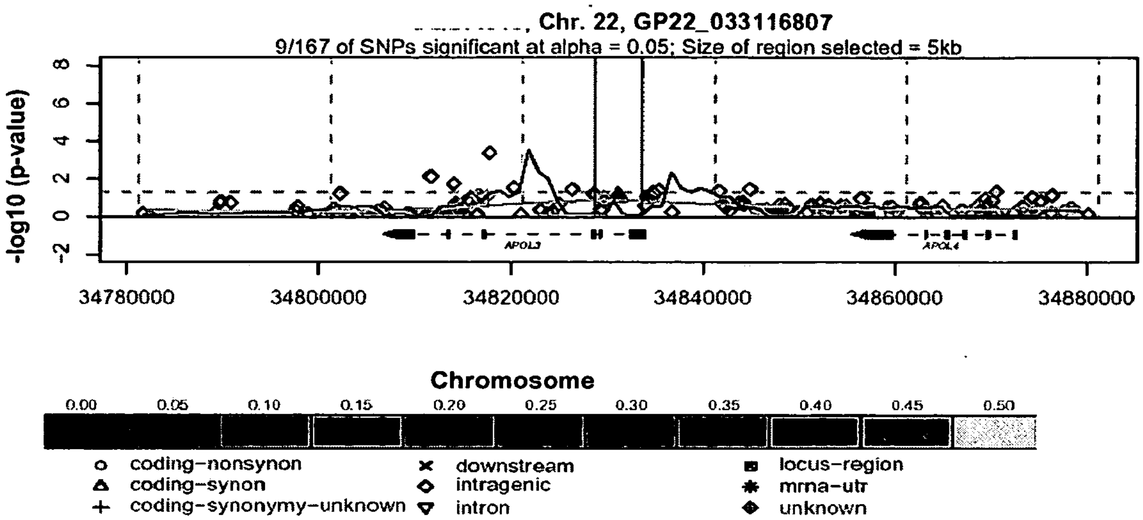
40. The composition of claim 38, wherein the RNA molecule is a short inhibitory RNA molecule.

Abstract of the Disclosure

Provided herein are methods for identifying a risk of osteoarthritis in a subject, reagents and kits for carrying out the methods, methods for identifying candidate therapeutics for treating osteoarthritis, and therapeutic and preventative methods applicable to osteoarthritis. These embodiments are based upon an analysis of polymorphic variations in nucleotide sequences within the human genome.

FIGURE 1

APOL3 – DISCOVERY P-VALUES (female only)



Application Data Sheet

Application Information

Application Type::	Provisional
Subject Matter::	Utility
Suggested Group Art Unit::	Not Yet Assigned
CD-ROM or CD-R?::	None
Sequence submission?::	None
Computer Readable Form (CRF)?::	No
Title::	METHODS FOR IDENTIFYING RISK OF OSTEOARTHRITIS AND TREATMENTS THEREOF
Attorney Docket Number::	524593009200
Request for Early Publication?::	No
Request for Non-Publication?::	No
Total Drawing Sheets?::	1
Small Entity?::	Yes
Petition included?::	No
Secrecy Order in Parent Appl.?::	No

Applicant Information

Applicant Authority Type::	Inventor
Primary Citizenship Country::	US
Status::	Full Capacity
Given Name::	Steven
Family Name::	MAH
City of Residence::	San Diego
State or Province of Residence::	CA
Country of Residence::	US
Street of mailing address::	12820 Via Nieve #74
City of mailing address::	San Diego
State or Province of mailing address::	CA

Postal or Zip Code of mailing address:: 92130

Applicant Authority Type:: Inventor
Primary Citizenship Country:: Germany
Status:: Full Capacity
Given Name:: Andreas
Family Name:: BRAUN
City of Residence:: San Diego
State or Province of Residence:: CA
Country of Residence:: US
Street of mailing address:: 3935 Lago Di Grata Circle
City of mailing address:: San Diego
State or Province of mailing address:: CA
Postal or Zip Code of mailing address:: 92130

Applicant Authority Type:: Inventor
Primary Citizenship Country:: Germany
Status:: Full Capacity
Given Name:: Stefan
Middle Name:: M.
Family Name:: KAMMERER
City of Residence:: San Diego
State or Province of Residence:: CA
Country of Residence:: US
Street of mailing address:: 3825 Elijah Court, Unit 334
City of mailing address:: San Diego
State or Province of mailing address:: CA
Postal or Zip Code of mailing address:: 92130

Applicant Authority Type:: Inventor
Primary Citizenship Country:: US
Status:: Full Capacity

Given Name::	Matthew
Middle Name::	Roberts
Family Name::	NELSON
City of Residence::	San Marcos
State or Province of Residence::	CA
Country of Residence::	US
Street of mailing address::	1250 Calle Prospero
City of mailing address::	San Marcos
State or Province of mailing address::	CA
Postal or Zip Code of mailing address::	92069

Applicant Authority Type::	Inventor
Primary Citizenship Country::	Sweden
Status::	Full Capacity
Given Name::	Rikard
Middle Name::	Henry
Family Name::	RENELAND
City of Residence::	San Diego
State or Province of Residence::	CA
Country of Residence::	US
Street of mailing address::	7555 Charmant Drive, #1114
City of mailing address::	San Diego
State or Province of mailing address::	CA
Postal or Zip Code of mailing address::	92122

Applicant Authority Type::	Inventor
Primary Citizenship Country::	United Kingdom
Status::	Full Capacity
Given Name::	Maria
Middle Name::	L.
Family Name::	LANGDOWN
City of Residence::	San Diego

State or Province of Residence::	CA
Country of Residence::	US
Street of mailing address::	3701 Yosemite Street
City of mailing address::	San Diego
State or Province of mailing address::	CA
Postal or Zip Code of mailing address::	92109

Correspondence Information

Correspondence Customer Number::	25225

Representative Information

Representative Customer Number::	25225